



Universidad Central de Venezuela

Facultad de Ciencias

Escuela de Computación

**PERCEPCIÓN SOCIAL DE LOS PROYECTOS
FINANCIADOS POR EL CONSEJO FEDERAL DE
GOBIERNO A PARTIR DE LOS CONTENIDOS
PUBLICADOS EN TWITTER.**

Trabajo Especial de Grado presentado ante la ilustre

Universidad Central de Venezuela por los

Br. Iraima Rodríguez

Br. Emmanuel Galeano

Tutores:

Prof. José Sosa

Prof(a). Joali Moreno

Caracas, Octubre del 2016

ACTA

Quienes suscriben, miembros del jurado designado por el Consejo de la Escuela de Computación para examinar el Trabajo Especial de Grado, presentado por los Bachilleres Iraima Rodríguez C.I.:19.659.513, y Emmanuel Galeano C.I.:19.658.075 titulado "PERCEPCIÓN SOCIAL DE LOS PROYECTOS FINANCIADOS POR EL CONSEJO FEDERAL DE GOBIERNO A PARTIR DE LOS CONTENIDOS PUBLICADOS EN TWITTER", aprobado por el Consejo de Escuela, a los fines de cumplir con el requisito legal para optar al título de Licenciado en Computación, dejan constancia de lo siguiente:

Leído el trabajo por cada uno de los miembros del jurado, se fijó el día 25 de Octubre de 2016, a las 8:00 AM, para que sus autores lo defendieran de forma pública, en el aula PAB, lo cual estos realizaron mediante una exposición oral de su contenido, y luego respondieron satisfactoriamente a las preguntas que le fueron formuladas por el jurado, todo ello conforme a lo dispuesto en la Ley de Universidades y demás normativas vigentes de la Universidad Central de Venezuela. Finalizada la defensa pública del Trabajo Especial de Grado, el jurado decidió aprobarlo.

En fe de lo cual se levanta la presente acta, en Caracas el día 25 de Octubre de 2016, dejándose también constancia de que actuó como Coordinador del jurado el Profesor Tutor José Sosa y Robinsón Rivas.

X 
Prof. Jose Sosa
Tutor

X 
Prof.(a)Joali Moreno
Tutor

X 
Prof. Robinson Rivas
Jurado

X 
Prof. Jesús Lares
Jurado



AGRADECIMIENTOS

Principalmente a Dios por permitirnos llegar hasta aquí dándonos la sabiduría y entendimiento para el desarrollo de este sistema, a nuestros familiares por siempre creer en nosotros y brindarnos su apoyo en todo momento.

A nuestros profesores José Sosa, Robinson Rivas y Joali Moreno por su constante trabajo, estando siempre disponible para ayudar con este proyecto. Fueron sin duda el factor decisivo para llevar el trabajo hasta donde está. A nuestros amigos y compañeros del Consejo Federal de Gobierno por su ayuda, además a todas las personas que hicieron esto posible.

Iraima Rodriguez y Emmanuel Galeano.

UNIVERSIDAD CENTRAL DE VENEZUELA.

FACULTAD DE CIENCIAS

ESCUELA DE COMPUTACIÓN

**PERCEPCIÓN SOCIAL DE LOS PROYECTOS FINANCIADOS POR EL
CONSEJO FEDERAL DE GOBIERNO A PARTIR DE LOS CONTENIDOS
PUBLICADOS EN TWITTER.**

Autores: Br. Iraima Rodriguez

Br. Emmanuel Galeano

Tutores: Prof. Jose Sosa

Prof. Joali Moreno

Fecha: Caracas, Noviembre de 2016

RESUMEN

El presente trabajo especial de grado expone el desarrollo de un sistema web, el cual fue elaborado con la finalidad de realizar estudios de percepción social de los proyectos del Consejo Federal de Gobierno el cual es el órgano encargado de la planificación y coordinación de políticas y acciones para el desarrollo del proceso de descentralización y transferencia de Competencias del Poder Nacional a los estados y municipios, a partir del análisis de los contenidos publicados en Twitter. En el desarrollo de este sistema, se aplicó una adaptación de la metodología AdHoc mediante la metodología ágil Programación extrema (XP) y la metodología CRISP-DM. Se utilizaron diversas herramientas de trabajo para el desarrollo de los componentes del sistema, como PHP, R, RStudio, y el API de Twitter que permite extraer información (tweets), los cuales se encuentran almacenados en una base de datos central, construida utilizando el gestor de base de datos MongoDB.

PALABRAS CLAVE: Consejo Federal de Gobierno, Twitter, percepción social

ÍNDICE

INTRODUCCIÓN	i
CAPÍTULO 1 PROBLEMA DE INVESTIGACION	1
1.1. PLANTEAMIENTO DEL PROBLEMA	1
1.2. JUSTIFICACIÓN	2
1.2.1. ¿POR QUÉ ES UN PROBLEMA?	2
1.2.2. ¿PARA QUIÉN ES UN PROBLEMA	2
1.2.3. ¿DESDE CUÁNDO ES UN PROBLEMA	2
1.3. OBJETIVOS	3
1.3.1. OBJETIVOS GENERALES	3
1.3.2. OBJETIVOS ESPECÍFICOS	3
1.4. ANTECEDENTES	3
1.4.1. CONSEJO FEDERAL DE GOBIERNO	3
1.5. ALCANCE	4
CAPÍTULO 2 MARCO CONCEPTUAL	
2.1. TECNOLOGÍAS DISPONIBLES	5
2.1.1. REDES SOCIALES DE MICROBLOGGIN: TWITTER	5
2.2. INTERFACES DE PROGRAMACIÓN DE APLICACIONES (API) DE TWITTER	5
2.2.1. RESTFUL	6
2.2.2. STREAMING API	6
2.3. BASES DE DATOS NOSQL	7
2.3.1. BASES DE DATOS NOSQL ORIENTADAS A CLAVE-VALOR	8
2.3.2. BASES DE DATOS NOSQL ORIENTADAS A COLUMNAS	8
2.3.3. BASES DE DATOS NOSQL ORIENTADAS A DOCUMENTOS	8
2.3.4. BASES DE DATOS NOSQL ORIENTADAS A GRAFOS	8
2.4. ANÁLISIS DE SENTIMIENTOS	9
2.5 MINERÍA DE TEXTO	10
2.6. MINERÍA DE REDES SOCIALES	11
2.7. SISTEMAS EXISTENTES DISPONIBLES	12

2.7.1. EXTRACCIÓN, ANÁLISIS Y VISUALIZACIÓN DE INFORMACIÓN SOCIAL DESDE TWITTER	12
2.7.2. FILTRADO Y DETECCIÓN DE TÓPICOS BASADO EN LA ENTIDAD PARA LA MONITORIZACIÓN DE LA REPUTACIÓN ONLINE EN TWITTER	13
2.7.3. USANDO TÉCNICAS DE IR PARA ANÁLISIS DE SENTIMIENTOS BASADOS EN TÓPICOS A TRAVÉS DE MODELOS DE DIVERGENCIA	16
2.7.4. APLICACIONES QUE REALIZAN GESTIÓN Y ANÁLISIS DE REDES SOCIALES (CYFE)	16
CAPÍTULO 3 MÉTODO DE DESARROLLO	
3.1. METODOLOGÍAS	22
3.1.1. CRISP-DM	22
3.1.2. METODOS AGILES	24
3.2. HERRAMIENTAS A UTILIZAR	30
CAPÍTULO 4 DESARROLLO DE LA SOLUCIÓN	
4.1. ARQUITECTURA DE LA SOLUCIÓN	32
4.2. ANÁLISIS Y DISEÑO DE LA SOLUCIÓN	35
4.3. DESARROLLO	53
4.3.1. ANÁLISIS EXPLORATORIO DE LOS DATOS	53
4.3.2. TRANSFORMACIÓN DE LA DATA	54
4.3.4. RESULTADOS	55
4.4 PRUEBAS	61
CAPÍTULO 5 CONCLUSIONES	
5.1. CONCLUSION	77
5.2. CONTRIBUCIÓN	79
5.3. RECOMENDACIONES	79
5.3. TRABAJOS FUTUROS	79
5.4. REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES	80

ÍNDICE DE FIGURAS

Figura1. Página web de inicio de Cyfe.	17
Figura2. Características de Cyfe en su página web de inicio.	18
Figura3. Widgets de Cyfe en su página web de inicio.	18
Figura4. Blog de Cyfe.	19
Figura4.1. Blog de Cyfe.	19
Figura 5. Versión Premium de Cyfe	20
Figura 6. Inicio de sesión de Cyfe.	20
Figura 7. Registro de Cyfe.	20
Figura 8. Metodologías utilizadas en Minería de Datos.	22
Figura 9. Diagrama de procesos que muestra la relación entre las diferentes fases de CRISP-DM.	23
Figura10.Ciclo de desarrollo de Programación Extrema (XP)	28
Figura 11. Arquitectura de la solución	35
Figura 12. Modelo de Base de Datos.	41
Figura 13. Diagrama de casos de uso Analista	42
Figura 13.1. Diagrama de casos de uso Administrador	43
Figura 14. Principal.	44
Figura 15. Formato del archivo de búsquedas.	45
Figura 16. Cargar búsquedas	45
Figura 17. Gestionar búsquedas.	46
Figura 18. Formato del Archivo de limpieza.	47
Figura 19. Mostrar archivo de limpieza.	47
Figura 19.1.Detalle archivo de limpieza	48
Figura 19.2. Modificar archivo de limpieza	48
Figura 20. Detalle Palabras positivas.	49
Figura 21. Mostrar Palabras negativas	49
Figura 21.1. Detalle Palabras negativas	50
Figura 22. Resultados por Nube de Palabras	50
Figura 23. Resultados por Histograma de polaridad	51
Figura 24. Vista Principal Administrador.	51

Figura 25 Modificar archivo de limpieza Administrador.	52
Figura 26 Modificar Palabras por defecto Administrador.	52
Figura 27 Modificar Palabras Negativas por Defecto Administrador.	53
Figura 28 Agregar, Modificar y Eliminar usuarios.	53
Figura 29. Dataframe	54
Figura 30. Tweets en formato JSON	55
Figura 31. Nubes de palabras	56
Figura 32. Histograma de polaridad	57
Figura 33. Reporte nube de palabras	58
Figura 33.1. Reporte nube de palabras	59
Figura 34. Reporte Histograma de Polaridad	60
Figura 34.1. Reporte de Histograma de Polaridad	61
Figura 35. Prueba1 Nube de Palabra	62
Figura 36. Prueba1 Reporte Nube	63
Figura 37. Prueba1 Histograma	64
Figura 38. Prueba1 Reporte Histograma	65
Figura 39. Prueba1 Texto Plano	66
Figura 40. Prueba2 Nube de Palabras.	67
Figura 41. Prueba2 Reporte Nube.	68
Figura 42. Prueba2 Histograma de Polaridad	69
Figura 43. Prueba2 Reporte Histograma de Polaridad	70
Figura 44. Prueba2 Texto Plano.	71
Figura 45. Prueba3 Nube de Palabra	72
Figura 46. Prueba3 Reporte Nube	73
Figura 47 Prueba3 Histograma de Polaridad	74
Figura 48. Prueba3 Reporte Histograma	75
Figura 49. Prueba3 Texto plano	76

INTRODUCCIÓN

El Consejo Federal de Gobierno, es el órgano encargado de la planificación y coordinación de políticas y acciones para el desarrollo del proceso de descentralización y transferencia de Competencia del Poder Nacional a los estados y municipios.

El presente trabajo especial de grado se propone el desarrollo de un sistema web, que permita realizar análisis de sentimientos por medio de los tweets extraídos y así conocer las opiniones que tienen las personas beneficiadas por los proyectos financiados por el Consejo Federal de Gobierno y el impacto que ha traído dicho proyecto, mediante nubes de palabras, reportes, y gráfico de polaridad.

En las investigaciones realizadas previamente en nuestro seminario se encontraron algunos sistemas y técnicas disponibles, los cuales realizan extracción, análisis y visualización de información social desde Twitter, asociación por palabras claves, detección de tópicos y análisis de sentimientos, obtener información relacionada a la percepción social de las empresas, que se utilizaran como base teórica para el desarrollo de esta investigación.

En el Capítulo 1 se plantea el problema, se explica desde cuándo es un problema y para quién es un problema, posteriormente se plantean los objetivos para atacar dicho problema.

En el Capítulo 2 se define las tecnologías disponibles y sistemas existentes disponibles, y se exponen cuáles fueron las tecnologías usadas para llevar a cabo el sistema, en el Capítulo 3 se trata todo lo referente a la metodología AdHoc adaptando las metodologías CRISP-DM y la metodología Ágil programación extrema (XP) . En el Capítulo 4, se define formalmente el desarrollo de la aplicación y los resultados obtenidos.

CAPITULO 1

PROBLEMA DE INVESTIGACION

1.1. PLANTEAMIENTO DEL PROBLEMA

El Consejo Federal de Gobierno cuenta con la Oficina de Seguimiento y Evaluación de Políticas Públicas, la cual se encarga de controlar, seguir y sistematizar los datos históricos y contingentes sobre el comportamiento de la ejecución de proyectos de inversión aprobados por el Fondo de Compensación Interterritorial (FCI).

Para que esta labor sea llevada a cabo, la oficina cuenta con varios sistemas y técnicas, que le permite controlar y seguir la ejecución de los proyectos aprobados por el Fondo de Compensación Interterritorial. Entre esas técnicas que utilizan están:

- Realizar visitas
- Verificación de rendición, los cuales realizan a través de índices de gestión
- Encuestas vías telefónicas

Desde hace un tiempo el Consejo Federal de Gobierno desea complementar dichas técnicas y sistemas que tienen para realizar seguimiento, y se crea la necesidad de conocer el nivel de satisfacción de las personas beneficiadas, por los proyecto de inversión aprobados.

Y es por esta necesidad que surge la primera pregunta por parte de ellos: ¿Cómo extraer nuevas variables para la toma de decisiones por medio de redes sociales específicamente Twitter?

Ellos necesitan una aplicación que les permita realizar análisis de percepción en redes sociales. Donde las personas puedan expresarse libremente.

En las redes sociales las personas tienen libertad para realizar publicaciones de diferentes índoles, permitiendo así expresar sus ideas, puntos de vista y sentimientos sin ninguna limitación, pueden plantear todo tipo de inquietud con plena confianza, esto permite la extracción de información importante de las mismas. En las redes sociales, particularmente en Twitter, la información no sigue un esquema determinado, ya que puede tener distintos formatos de imágenes, vídeos, audio y textos. Se busca extraer nueva información del texto contenido en los tweets que sea relevante para la toma de decisiones del Fondo de Compensación Interterritorial.

Para lograr esto se investigaron las diferentes tecnologías existentes, entre ellas están el análisis de sentimientos, minería de texto, minería de redes sociales y agrupación.

De la problemática planteada se constituye la siguiente pregunta: ¿Cómo conocer el nivel de satisfacción de las personas, con respecto a los proyectos financiados por el Consejo Federal de Gobiernos utilizando la red social Twitter?

1.2. JUSTIFICACIÓN

1.2.1. ¿Por qué es un problema?

Es un problema porque el Consejo Federal de Gobierno, ya que no cuentan con un sistema que cubra con la necesidad que se les presente, la cual es, conocer el impacto que los proyectos aprobados han causado tanto a nivel personal como comunitario viendo así las necesidades, quejas, o nivel de satisfacción que las personas puedan tener. Donde se expresarse sin ninguna limitación y así el Consejo Federal de Gobierno pueda ver la realidad que han causado los proyectos financiados y a su vez observar las necesidades que puedan existir en comunidades. Permitiendo reforzar los sistemas y técnicas existentes.

1.2.2. ¿Para quién es un problema?

Es un problema para el Consejo Federal de Gobierno ya que no cuentan con un sistema que les permita realizar análisis de percepción social.

1.2.3. ¿Desde cuándo es un problema?

Desde el primer momento que el Consejo Federal de Gobierno se propone disminuir los desequilibrios en los Estados a través de proyectos de inversión para Entidades Político Territoriales y Organizaciones bases del poder popular.

Con el fin de dar una solución al problema planteado, se establecen los siguientes objetivos para este trabajo especial de grado.

1.3. OBJETIVOS

1.3.1. Objetivo General

Desarrollar una aplicación web para realizar análisis de percepción social por medio de twitter para su uso en el Consejo Federal de Gobierno.

1.3.2. Objetivos Específicos

- Levantamiento de requerimientos
- Diseñar la interfaz: Diseñar una interfaz para el usuario final donde pueda realizar todos los procesos de búsqueda, limpieza, análisis, y resultado de una forma intuitiva.
- Crear roles y usuarios: Crear dos tipos de roles, administrador y analista los cuales se asignaran a los distintos usuarios.
- Realizar extracción, limpieza y procesamiento: Realizar extracción de los Twitter desde el API de Twitter que luego serán limpiados quitando los artículos, acentos, número, iconos entre otros para ser procesados.
- Mostrar resultados de los procesamientos de la información recopilada: Se mostraran los resultados de la extracción, limpieza y análisis que se realizó anteriormente en forma de nubes de palabras, histograma de polaridad, texto plano y reportes.

1.4. ANTECEDENTES

1.4.1. Consejo Federal de Gobierno

El Consejo Federal de Gobierno tiene como función establecer los lineamientos en materia de descentralización entre las entidades político territoriales y hacia las organizaciones de base del Poder Popular, así como para el estudio, planificación y creación de los Distritos Motores de Desarrollo, a los fines de apoyar especialmente la organización popular y el desarrollo de obras y servicios esenciales en las regiones y comunidades de menor desarrollo relativo, siempre enmarcado en el Plan de Desarrollo Económico y Social de la Nación. [1].

1.5. ALCANCE

El desarrollo de este proyecto viene dado por la construcción de una aplicación web enfocada al estudio de la percepción social de los proyectos del Consejo Federal de Gobierno a partir del análisis de los contenidos publicados en Twitter. Esta aplicación permitirá:

- Realizar búsquedas por cuadro de texto.
- Permite al usuario realizar búsquedas por medio de un archivo.
- Realizar estadística básica mediante reportes.
- Hacer análisis de tópicos por medio de nubes de palabras.
- Hacer análisis de sentimientos mediante gráficos de polaridad.
- Área de configuración que permite modificar diccionarios, archivo de limpieza y parámetros de búsqueda.
- Dos tipos de usuarios, tales como:
 - **Analista:** podrá acceder a todas las funcionalidades del sistema.
 - **Administrador:** podrá crear usuarios y se encargara del área de configuración que permite modificar y eliminar archivo de limpieza y parámetros de búsqueda.

En este trabajo especial de grado no se realizara análisis en tiempo real, ni se maneja las fechas ni ubicación geográfica con respecto a los tweets, tampoco se realizara análisis de grupos.

CAPITULO 2

MARCO CONCEPTUAL

2.1. TECNOLOGÍAS DISPONIBLES

En la presente sección se muestran ejemplos de aplicaciones y técnicas que utilizan diferentes empresas para realizar gestión y análisis, por medio de las redes sociales en búsqueda de sus objetivos, así como también, obtener información relacionada a la percepción social de las empresas.

2.1.1. Redes sociales de microbloggin: Twitter

Dentro de los sistemas de microblogging, Twitter es uno de los que ha tenido un mayor auge al menos en el mundo occidental, muchos analistas consideran que Twitter es mucho más adecuado para el análisis del flujo de noticias y opiniones de todo tipo, lo que lo hace idóneo para el análisis de sentimiento, estudios de reputación, entre otros. “Un microblogging es una nueva forma de comunicación donde los usuarios pueden escribir su estado actual en mensajes cortos, distribuidos automáticamente por mensajes, teléfonos móviles, correos o vía web. Twitter es un popular microblogging lanzado en octubre de 2006.” [2].

En nuestro trabajo especial de grado estamos utilizando el microbloggin de Twitter como fuente de información, de la cual extraemos los datos o información con los cuales vamos a trabajar.

2.2. INTERFACES DE PROGRAMACIÓN DE APLICACIONES (API) DE TWITTER

Twitter es una de las redes sociales de mayor crecimiento, basada en el concepto de "microblogging", que permite a los usuarios postear mensajes de una longitud reducida en números de caracteres. A través de las API de Twitter cualquiera puede crear aplicaciones que comuniquen con el servicio de la mencionada red social. Un API o interfaz de programación de aplicaciones, es un conjunto de funciones y procedimientos, para ser utilizados por otro software, brindando un nivel de abstracción al programador. La API de Twitter tiene dos formas de trabajo, RESTful y Streaming.

2.2.1. Restful

RESTful API ofrece acceso a los datos almacenados en el núcleo de Twitter. Proporciona acceso para leer y escribir datos en Twitter, acceso a la información de los perfiles de usuarios y a sus seguidores. Una limitante de la RESTful API es el tiempo límite de la ventana, el cual es de 15 minutos por solicitud, las solicitudes GET tienen disponibles en este tipo de API dos formas de consultar, una donde se harán 150 solicitudes y otra donde se harán 180 solicitudes cada 15 minutos.[3].

La RESTful API identifica a los usuarios y aplicaciones utilizando la autenticación OAuth, en este tipo de autenticación la aplicación realiza peticiones al API sin involucrar al usuario, las solicitudes al API están limitadas pero esto ocurre por los métodos del API, no es un límite por usuario. En la documentación existen dos límites para la frecuencia utilizando esta forma de autenticación, uno es por usuario (para la autenticación de aplicaciones de usuario) y la otra por la aplicación (por esta forma de autenticación solo por aplicación). La desventaja de OAuth es que no todos los métodos del API soportan la autenticación solo por aplicación.[4]

Las ventajas de la autenticación OAuth son:

- Los usuarios no están obligados a compartir sus claves con aplicaciones de terceros, esto incrementa la seguridad.
- Una gran cantidad de bibliotecas son compatibles con Twitter utilizando OAuth.

2.2.2. Streaming API

La Streaming API permite un acceso con poca latencia a la información de Twitter, proporciona un subconjunto del flujo total de tweets en tiempo real. La ventaja del Streaming API es que no tiene un límite de peticiones al servidor, pero retorna tweets en tiempo real, lo cual no permite realizar análisis histórico.[5].

Las diferencias entre RESTful API y Streaming API son:

- Para conectarse con Streaming API es necesario mantener abierta una conexión HTTP persistente, con RESTful se realiza la conexión con OAuth.
- RESTful API tiene una limitación de 150 peticiones/hora por usuario o por IP si la conexión no está autenticada, Streaming API no tiene esta limitante.

- Streaming API obtiene tweets en tiempo real, RESTful obtiene tweets desde el núcleo de Twitter permitiendo realizar análisis histórico de esta información.

Para el desarrollo de nuestra aplicación decidimos usar RESTful API ya que nos permite realizar análisis históricos sobre los datos de Twitter y con ello monitoriar el estado de los proyectos en base a las publicaciones realizadas por los usuarios a través del tiempo.

2.3. BASES DE DATOS NOSQL

Para realizar análisis histórico de la información contenida en una red social, se deben almacenar los datos desde cierta fecha hasta otra fecha (rango de tiempo), para luego ser almacenada en algún lugar, en la mayoría de los casos la información es voluminosa.

Para esta tarea se utilizan hoy en día Bases de Datos no relacionales las cuales facilitan la indexación y la búsqueda. Gracias a estos avances se pueden manejar enormes cantidades de datos.

Algunas ventajas de las bases de datos NoSQL son:

- Manejan grandes cantidades de información estructurada, semi estructurada y no estructurada.
- Tienen programación orientada a objetos, lo cual las hace flexibles y de fáciles de usar.
- Están geográficamente distribuidas, permitiendo un crecimiento horizontal.
- Son software libre.[6].

Algunas limitaciones de las bases de datos NoSQL:

- A menudo no cumplen atomicidad, consistencia y durabilidad.
- Tiene consultas más limitadas que las bases de datos relacionales y requieren trasladar complejidad a la aplicación.
- Aún falta madurez. [7].

2.3.1. Base de datos NOSQL orientada a clave-valor

Es un tipo de base de datos NoSQL que consiste en un mapa o diccionario donde se almacenan valores que están asociados a una clave con la cual se pueden acceder rápidamente. Este tipo de base de datos tiene las ventajas de ser simple, escalable, controla grandes volúmenes de información, maneja distintos tipos de datos, gran velocidad para agregar y acceder a la información.

Tiene limitaciones en cuanto a las funciones analíticas, consultas complejas y disparadores.[8].

2.3.2. Base de datos NOSQL orientada a columnas

Este tipo de base de datos NoSQL nació gracias a la aplicación BigTable de Google y la implementación de Cassandra de Apache. Se divide la información en un conjunto de columnas, esta organización aumenta el rendimiento de las consultas de agregación (count, sum, avg, min, max), carga rápida de grandes volúmenes de datos y gran accesibilidad a muchas herramientas de inteligencia de negocios. [8].

Una desventaja de este almacenamiento es el costo en reconstrucción de las tuplas y el incremento en costo de las inserciones.

2.3.3. Base de datos NOSQL orientada a documentos

Las bases de datos documentales son consideradas como el siguiente paso de las bases de datos orientadas a clave-valor, ya que almacenan pares clave-valor en estructuras más complejas llamadas documentos, al no tener un esquema estricto para definir los documentos su uso se simplifica y mejoran su sistema de indexado utilizando árboles binarios.[9].

Generalmente utilizan la estructura simple del JSON o XML con una clave única para cada registro, son las bases de datos NoSQL más versátiles.

2.3.4. Base de datos NOSQL orientada a grafos

Inspiradas en la teoría de grafos los datos son representados utilizando nodos y sus relaciones como aristas que unen a los nodos. Un grafo es una estructura flexible y capaz

de integrarse fácilmente con aplicaciones orientadas a objetos. Este tipo de base de datos es recomendada para utilizarse en conjunto con redes sociales.[10].

Las fortalezas de una base de datos de este tipo es su flexibilidad dado que se acomoda a las necesidades del negocio, optimiza el rendimiento y su gran escalabilidad. Tiene limitaciones al momento de buscar información en nodos de diferentes maquinas, esto puede generar un costo en tiempo y que requiere de tiempo para que los desarrolladores aprendan a utilizarla. En nuestro trabajo especial de grado fue elegida la base de datos MongoDB por ser la base de datos líder y permite a las empresas ser más ágiles y escalables y por estar especializada para el manejo de datos del tipo texto.

Para nuestro modelo de base de datos elegimos el tipo documental ya que es más sencillo de utilizar, versátil y no es restrictivo. Además es con el que tenemos más experiencia y existe una gran cantidad de documentación.

2.4. ANÁLISIS DE SENTIMIENTOS

En la última década, el análisis de sentimientos, ha despertado un creciente interés. Resulta un gran reto para las tecnología del lenguaje, ya que obtener buenos resultados es mucho más difícil de lo que muchos creen.

La tarea de clasificar automáticamente un texto escrito en un lenguaje natural en un sentimiento positivo o negativo, opinión o subjetividad, es a veces tan complicada que incluso es difícil poner de acuerdo a diferentes anotadores humanos sobre la clasificación a asignar a un texto dado. La interpretación personal de un individuo es diferente de la de los demás, y además se ve afectada por factores culturales y experiencias propias de cada persona. Y la tarea es aún más difícil cuanto más corto sea el texto, y peor escrito esté, como es el caso de los mensajes en redes sociales como Twitter o Facebook.

“El análisis de sentimientos es un proceso en el cual intentamos predecir el tipo de sentimiento que una persona pudo sentir o intento expresar al momento de escribir cierta información, en este caso mediante una publicación en la red social Twitter. El objetivo de este trabajo es el de determinar si el usuario de twitter realizo una publicación optimista, pesimista o ninguna de las anteriores.”[11]. El análisis de sentimientos es un proceso en el cual se determina la polaridad del usuario (a favor, en contra o neutra)

sobre un tópico en una o varias publicaciones. Con el análisis de sentimientos se desea obtener información útil para la toma de decisiones de negocio, por ejemplo, cambiar la estrategia con la cual se promociona un candidato para un cargo político. Otro ejemplo de decisión de negocio basada en análisis de sentimientos es cambiar aspectos o características de un producto o sacar del mercado dicho producto para evitar más pérdidas.

Algunos casos donde se realiza el análisis de sentimientos son:

- Determinar la apreciación de una personalidad o producto en la comunidad.
- Determinar la reputación de una institución pública o privada.

Entre sus dificultades están:

- Distinguir expresiones sarcásticas.
- La complejidad de la estructura del lenguaje natural.

Lo implementamos en nuestro trabajo especial de grado analizando la polaridad de los comentarios relacionados a los proyectos financiados por el Consejo Federal de Gobierno, visualizando así el nivel de satisfacción de las personas con respecto a los proyectos antes mencionados.

2.5 MINERÍA DE TEXTO

En el mundo de la minería de datos todo proceso de análisis comienza con la construcción y extracción de un conjunto de datos que se llama vista minable. La vista minable generalmente viene del manejo de datos de una o varias bases de datos estructuradas, cuando queremos extraer información de textos, se tiene el problema de que los datos no son estructurados y para procesarlos se necesita de una cantidad de técnicas que se han constituido en un área de estudio propia, llamada minería de textos (text mining).

La minería de texto se refiere al proceso de derivar información nueva de documentos o textos que no está literalmente escrita. Utilizando algoritmos para la extracción y el análisis de datos de texto. El propósito de Minería de Texto es procesar la información no estructurada de los documentos o textos, para extraer patrones, tendencias, desviaciones o índices numéricos significativos del texto, los cuales pueden

ser utilizados por algoritmos estadísticos y por máquinas de aprendizaje (machine learning). Se pueden contabilizar las palabras y los grupos de palabras que se utilizan en los documentos; con esto se pueden analizar documentos y determinar similitudes entre ellos o cómo se relacionan con otras variables de interés, incluso crear resúmenes de los textos originales.[12].

Toda esta información se utilizara para la toma de decisiones, como determinar plagios en artículos científicos, clasificar grandes cantidades de documentos por tópicos de forma automática o permitirle a una empresa conocer mejor a sus clientes dadas sus preferencias y hábitos.

Algunas de las limitaciones de la Minería de Texto son:

- El acceso a todo el documento o texto ya que pueden estar restringidos por el autor.
- La cantidad de recursos utilizados en casos de grandes cantidades de documentos, como memoria y procesadores.
- Las API's no están estandarizadas para realizar las búsquedas de los documentos o textos.
- La gran cantidad de documentos y textos publicados a los cuales se les aplicaran los procesos de análisis de texto.
- Existen textos o documentos en los cuales se combinan diferentes idiomas, esto complica el proceso de análisis sobre los mismos.

En nuestro trabajo especial de grado aplicamos la minería de texto cuando extraemos las palabras que consideramos importantes para el proceso de análisis y desechamos aquellas que no aporten información relevante tales como artículos, números, enlaces, entre otras.

2.6. MINERÍA DE REDES SOCIALES

Las redes sociales proveen de mucha información la cual puede tener gran valor en algunos casos, por eso se utilizan los principios de la minería de datos aplicados a las redes sociales, con el fin de interpretar dicha información no estructurada y aprovecharla al máximo. Esta tecnología se conoce como Minera de Redes Sociales (Social mining). “Es el proceso de representar, analizar y extraer patrones procesables a partir de datos de

medios sociales.”[13].

Es un campo multidisciplinario, abarcando técnicas de la ciencia de la computación, la minería de datos, aprendizaje automático, análisis de redes sociales, ciencia de las redes, la sociología, la etnografía, la estadística, optimización, y matemáticas.[13].

Minería de Redes Sociales representa el mundo virtual de los medios sociales en una forma procesable para la computadora, capaz de ser medida y diseña modelos que pueden ayudar a entender sus interacciones. Además provee las herramientas necesarias para explotar el mundo de los patrones interesantes, analizar la difusión de la información, estudiar la influencia, provee recomendaciones efectivas para la toma de decisiones y analiza el comportamiento insólito en los medios sociales.

Minería de Redes Sociales se ve limitado por:

- La variedad de tipos de datos que son publicados en las redes sociales, imágenes, textos, vídeos, entre otros.
- Las fuentes de datos aparecen y desaparecen diariamente (redes sociales), por lo cual no tienen un comportamiento estable y dificulta mantenerlas identificadas para realizar el análisis.
- Es difícil determinar el universo y más aún la muestra adecuada con la cual obtener un resultado válido [14].

En nuestro trabajo especial de grado aplicamos la minería de redes sociales cuando mediante el API de Twitter restful nos conectamos a Twitter y extraemos los tweets que contengan las palabras o conjuntos de palabras relacionadas a los proyectos.

2.7. SISTEMAS EXISTENTES DISPONIBLES

En este segmento se explicaran algunos sistemas existentes, los cuales realizan extracción, análisis y visualización de información social desde Twitter, asociación por palabras claves, detección de tópicos y análisis de sentimientos, que se utilizaran como base teórica para el desarrollo de esta investigación.

Entre estos sistemas tenemos:

2.7.1. Extracción, análisis y visualización de información social desde twitter.

Hoy en día, cada vez tiene más importancia que el contenido de la web sea accesible

en el mismo momento de su creación. Al mismo tiempo, Twitter es una red social ampliamente utilizada para acceder a información en tiempo real ya que la gran mayoría de su contenido es accesible de forma pública. El objetivo de este proyecto es la extracción y análisis de información accesible a través de Twitter, así como la investigación de las posibilidades existentes para su procesamiento y posterior visualización. En este proyecto se hace una revisión tanto de artículos de investigación como de servicios relacionados con el uso de información que provee Twitter, seguida de la definición de un marco teórico que clasifique toda esa información. Se presenta el diseño de un sistema orientado en la extracción y procesamiento de información obtenida desde Twitter en español. Se han determinado tres estrategias de generación de información: la detección de género de los usuarios, la categorización de tweets por contenido el posicionamiento de tweets por áreas geográficas. Adicionalmente, el sistema ofrece a aplicaciones externas la posibilidad de acceder a la información generada.

Por último, se describe como ejemplo de uso una aplicación web que permite visualizar la información recogida y procesada por el sistema de diferentes formas. En ella se puede tanto interactuar con información en tiempo real como visualizar de forma gráfica la información almacenada. [16].

2.7.2. Filtrado y detección de tópicos basado en entidad para la monitorización de la reputación online en twitter.

Con el crecimiento de los medios sociales de comunicación en línea como Twitter (el servicio más popular de microblogging), los usuarios y consumidores han pasado a tener el control de lo que se dice acerca de una entidad (p.e., una compañía, un personaje público o una marca) en la Web. Este fenómeno ha creado la necesidad de monitorizar la reputación de dichas entidades en línea. En este ámbito, es esperable un aumento de la demanda de software de minería de textos para la monitorización de la reputación en línea (en inglés, Online Reputation Monitoring): herramientas automáticas que ayudan a procesar, analizar y agregar grandes flujos de menciones acerca de una compañía, organización o personaje público. A pesar de la gran variedad de herramientas disponibles en el mercado, no existe aún un marco de evaluación estándar (es decir, un

conjunto de tareas bien definidas, métricas de evaluación y colecciones reutilizables ampliamente aceptados) que permita abordar este problema desde un punto de vista científico.

En un marco de esfuerzo colectivo para identificar y formalizar los principales desafíos en el proceso de gestión de reputación en Twitter, hemos participado en la definición de tareas de acceso a la información, así como en la creación de colecciones de test (utilizadas en las campañas de evaluación WePS-3, RepLab 2012 y RepLab 2013) y hemos estudiado en profundidad dos de los desafíos identificados: filtrado de contenido no relevante (¿está relacionado un tweet dado con la entidad de interés?), modelado como una tarea de clasificación binaria, y detección de temas (¿qué se dice de la entidad en un flujo de tweets dado?), donde los sistemas deben agrupar los tweets en función de los temas tratados. En comparación con otros estudios sobre Twitter, nuestro problema se encuentra en su cola larga: salvando algunas excepciones, el volumen de información relacionado con una entidad dada (organización o compañía) en un determinado intervalo de tiempo es varios órdenes de magnitud más pequeño que los trending topics de Twitter, aumentando así su complejidad respecto a la identificación de los temas más populares en Twitter. En esta tesis nos basamos en tres conceptos para proponer distintas aproximaciones para abordar estas dos tareas: el uso de términos clave filtro (filter keywords), el uso de recursos externos (como Wikipedia, páginas web representativas de la entidad, etc.) y el uso de datos de entrenamiento específicos de la entidad (cuando éstos estén disponibles). Nuestros experimentos revelan que la noción de términos clave filtro (palabras que indican una alta probabilidad de que el tweet en el que aparecen esté relacionado o no con la entidad de interés) puede eficazmente ser utilizada para resolver la tarea de filtrado. En concreto, (a) la especificidad de un término con respecto al flujo de tweets de la entidad es un rasgo útil para identificar términos clave; y (b) la asociación entre el término y la página de la entidad en Wikipedia es útil para distinguir entre términos filtro positivos y negativos, especialmente cuando se calcula su valor medio teniendo en cuenta los términos más concurrentes. Además, estudiando la naturaleza de los términos filtro hemos llegado a la conclusión de que existe una brecha terminológica entre el vocabulario que caracteriza la entidad en Twitter y el vocabulario asociado a la entidad en su página principal,

Wikipedia o en la Web en general. Por otro lado, hemos hallado que, cuando se dispone de material de entrenamiento para la entidad en cuestión, es más efectivo el uso de un simple clasificador basado en bolsa de palabras. Existiendo suficientes datos de entrenamiento (unos 700 tweets por entidad), estos clasificadores pueden ser utilizados eficazmente para resolver la tarea de filtrado. Además, pueden utilizarse con éxito en un escenario de aprendizaje activo (active learning), en el que el sistema va actualizando su modelo de clasificación en función del flujo de anotaciones realizadas por el experto de reputación durante el proceso de monitorización. En este contexto, seleccionados los tweets en los que el clasificador tiene menos confianza (muestreo basado en márgenes) como aquellos que deben ser etiquetados por el experto, el coste de crear el conjunto inicial de entrenamiento puede llegar a reducirse en un 90% sólo inspeccionando el 10% de los datos de test. A diferencia de otras tareas de Procesamiento del Lenguaje Natural, el muestreo basado en márgenes funciona mejor que un muestreo aleatorio. Con respecto a la tarea de detección de temas, hemos considerado principalmente dos estrategias: la primera, inspirada en la noción de palabras término filtro, consiste en agrupar términos como un paso intermedio para la agrupación de tweets. La segunda, más exitosa, se basa en aprender una función de similitud entre pares de tweets a partir de datos previamente anotados, utilizando tanto rasgos basados en contenido como el resto de señales proporcionadas por Twitter; luego se aplica un algoritmo de agrupación sobre la función de similitud aprendida previamente. Nuestros experimentos revelan que (a) las señales Twitter pueden usarse para mejorar el proceso de detección de temas con respecto a utilizar sólo señales basadas en contenido; (b) aprender una función de similitud a partir de datos previamente anotados es una forma flexible y eficiente de introducir supervisión en el proceso de detección de temas. El rendimiento de nuestro mejor sistema es sustancialmente mejor que las aproximaciones del estado del arte, y se acerca al grado de acuerdo entre anotadores en las anotaciones de detección de temas incluidas en la colección RepLab 2013 (a nuestro conocimiento, la colección más grande para la monitorización de la reputación en línea). Una inspección cualitativa de los datos muestra que existen dos tipos de temas detectados por los expertos de reputación: alertas o incidentes de reputación (que normalmente sobresalen en el tiempo) y temas organizacionales (que, en cambio, suelen ser estables en el tiempo). Junto con nuestra

contribución para crear un marco estándar de evaluación para el estudio del problema de la monitorización de la reputación en línea desde una perspectiva científica, creemos que el resultado de nuestra investigación tiene implicaciones prácticas que pueden servir para beneficiar el desarrollo de herramientas semi-automáticas que asistan a los expertos en reputación en su trabajo diario de monitorización. [17].

2.7.3. Usando técnicas de IR para el análisis de sentimientos basado en tópicos a través de modelos de divergencia.

En este artículo se presenta el trabajo realizado para el Taller de Análisis de Sentimientos en la SEPLN. Este taller está enfocado al análisis de sentimientos en Twitter, tanto a nivel de tweet como a nivel de temática. Nuestra propuesta aborda la detección de sentimientos y temáticas mediante un sistema de Recuperación de Información (RI) basado en modelos del lenguaje. Se hace uso de la divergencia de Kullback-Liebler (KLD) para la generación tanto de los modelos de polaridad como de los modelos de temática que serán utilizados en el proceso de RI. Con el fin de mejorar la precisión de los resultados, se proponen varias aproximaciones centradas en llevar a cabo la obtención de los modelos del lenguaje considerando no sólo los contenidos textuales completos asociados a cada tweet sino, como alternativa, las entidades nombradas o los adjetivos detectados. Los resultados muestran como el uso tanto de entidades nombradas como de adjetivos en el modelo mejora los resultados de precisión obtenidos, indicando una mayor representatividad de éstos frente al uso de términos comunes. Los resultados generales son prometedores (5o y 4o posición en cada una de las tareas propuestas), lo que indica que una aproximación basada en RI y en modelos del lenguaje puede resultar una alternativa a otras propuestas más comunes en el estado del arte y centradas en la aplicación de técnicas clásicas de clasificación.[18].

2.7.4. Aplicaciones que realizan gestión y análisis sociales (CYFE).

Cyfe, es una aplicación que unifica las distintas métricas de una página web, un blog o redes sociales en un dashboard para realizar funciones tanto de gestión como de análisis. Algunas de sus características son:

- Generar estadísticas en tiempo real.

- Posibilidad de monitorear por departamento de la empresa, tal como, administración, atención al cliente y ventas.
- Facilidad de uso.
- Personalizado.
- Registra datos históricos.
- Ofrece seguridad para su uso con bases de datos y también utilizando Push API, cargar la información de tu base de datos en los dashboards.
- Capacidad de exportar los datos en diferentes formatos, como PNG, JPEG, PDF y CSV.
- Modo TV, con el cual se rotan automáticamente los dashboards en el monitor.
- Gran variedad de widgets para distintas redes sociales, como YouTube, Google Analytics, Twitter, entre otras.

En la figura 1 se observa la página de inicio de Cyfe, se muestra su lema “ ALL-In-One Dashboard”, algunos usuarios, obtener la versión premium, solicitar información, acceso al blog, opción de registrarse e iniciar sesión.



Figura1. Página web de inicio de Cyfe.

En la figura 2 se observa la continuación de la página de inicio de Cyfe, donde se muestran las características de la aplicación.

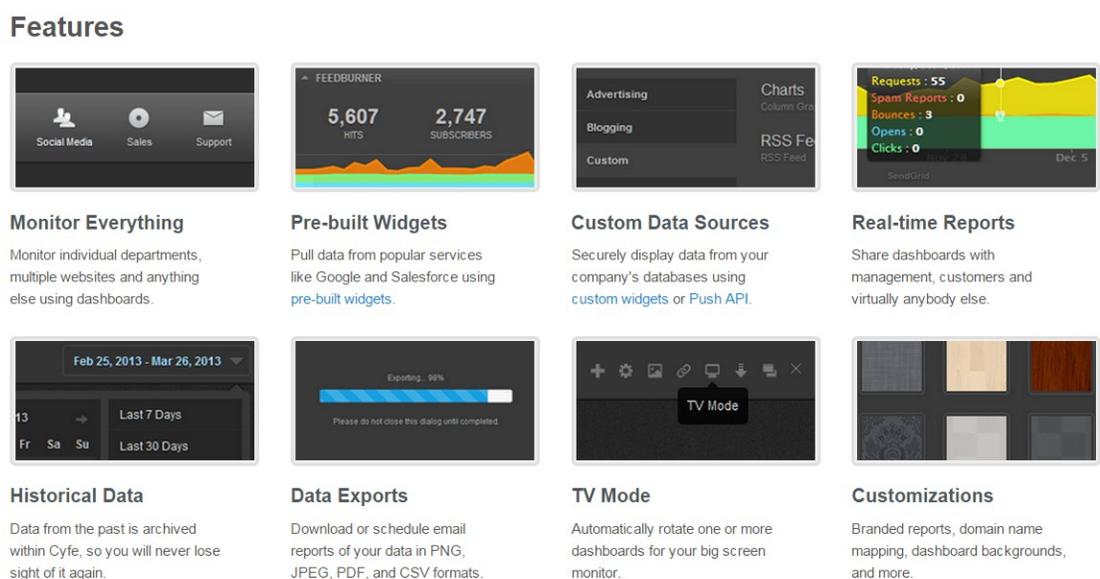


Figura2. Características de Cyfe en su página web de inicio.

En la figura 3 continua la página de inicio de Cyfe, con diferentes widgets disponibles en la aplicación.



Figura3. Widgets de Cyfe en su página web de inicio.

En la figura 4 y 4.1 se puede observar el blog que tiene la página de Cyfe, donde las personas suben sus opiniones, comentarios y experiencias con respecto a la aplicación.



Figura4. Blog de Cyfe.

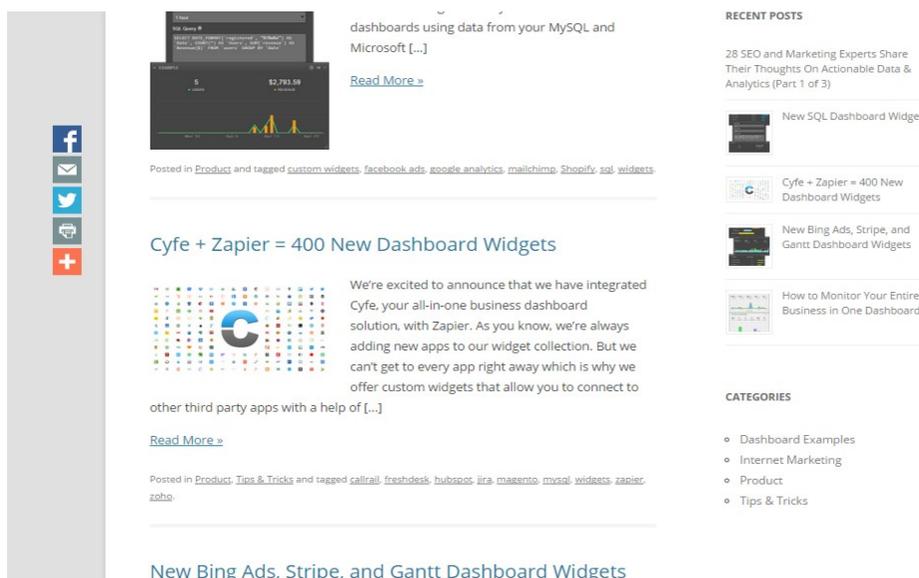


Figura4.1. Blog de Cyfe.

En la figura 5 se puede observar, las ventajas que trae descargar la versión premium

de Cyfe, tales como, obtener data histórica, exportar reportes en diferentes formatos, compartir dashboard de forma privada con miembros de un equipo, crear logo personalizado, compartir dashboards con cualquier persona en el mundo de solo lectura, personalizar cuenta con su propio nombre de dominio y rotar automáticamente los dashboards en el monitor además obtener más dashboards y widgets.



Figura 5. Versión Premium de Cyfe

En la figura 6 se puede observar, el inicio de sesión para ingresar a tu cuenta de Cyfe.



Figura 6. Inicio de sesión de Cyfe.

En la figura 7 se puede observar, el registro para crear una cuenta de Cyfe.

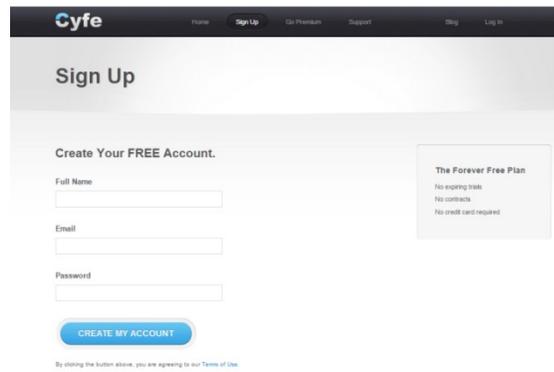


Figura 7. Registro de Cyfe.

En nuestro trabajo especial de grado investigamos una serie de artículos científicos para tener una base teórica que nos permita tener un apoyo referente a la aplicación que queremos desarrollar estos artículos reflejan sistemas que fueron implementados y que tienen características parecidas al sistemas que se está llevando a cabo. Particularmente Cyfe tiende a lo que queremos que es el análisis de sentimientos por medio de las redes sociales.

CAPÍTULO 3

MÉTODO DE DESARROLLO

3.1. METODOLOGÍAS

En nuestro trabajo especial de grado utilizamos la metodología AdHoc que es fusionar y adaptar diferentes metodologías según nuestra necesidad; en nuestro caso tomamos los tres primeros pasos de la metodología CRISP-DM tales como: comprensión del negocio, comprensión de datos y preparación de los datos los cuales serán explicados más adelante.

Y de la metodología XP tomamos los pasos diseño codificación y pruebas.

3.1.1 CRISP-DM

CRISP-DM o El Estándar para Procesos de Entre las Industrias para la Minería de Datos (Cross Industry Standard Process for Data Mining en inglés), es una metodología de desarrollo para los proyectos de Minería de Datos (así como también algunos proyectos de Ciencias de Datos y Datawarehousing) que posee gran participación y relevancia en el mercado de los proyectos de minería de datos como metodología de desarrollo, como se puede constatar en la figura 8. [19]

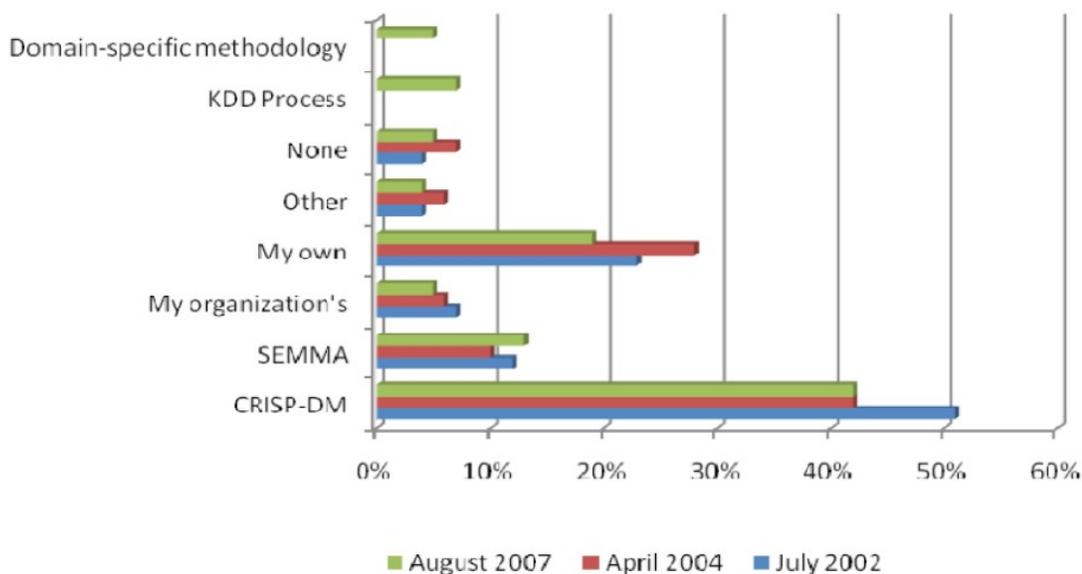


Figura 8. Metodologías utilizadas en Minería de Datos.

Como se puede ver en la figura 9. CRISP-DM está definido por 6 pasos iterativos.

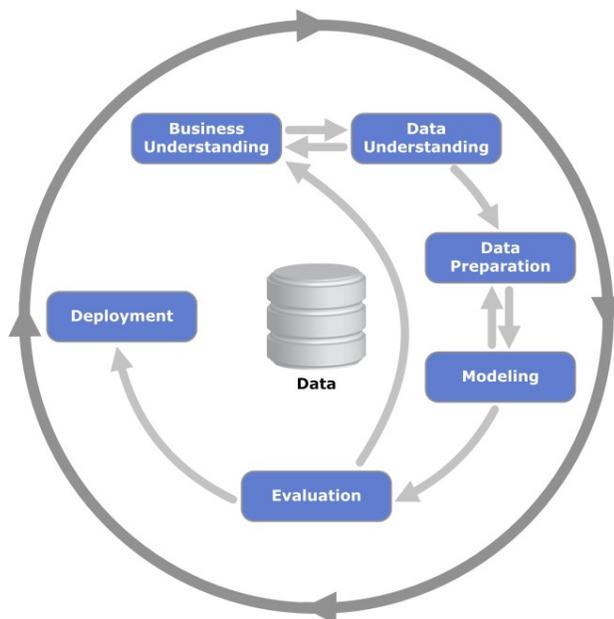


Figura 9. Diagrama de procesos que muestra la relación entre las diferentes fases de CRISP-DM.

Estos 6 pasos permiten el desarrollo de proyectos de minería de datos. A continuación se describen sus pasos:

- **Comprensión del negocio:** se establecen los objetivos y planes del proyecto de minería de datos atendiendo las necesidades y requisitos empresariales, determinando objetivos, requisitos, supuestos, restricciones, entre otros.
- **Comprensión de Datos:** esta fase se basa en comprender la data, sus fuentes, sus tipos, su creación, su utilidad, su calidad, entre otros. Esta fase permite saber de qué se dispone y permite identificar posibles puntos clave para encontrar relaciones entre los datos. Esta fase suele llamarse también fase de exploración de datos.
- **Preparación de datos:** Esta fase cubre, entre varias cosas, todo lo referido al proceso de ETL (Extraction, Transformation and Load o Extracción, Transformación y Carga en español). En esta fase se hace la búsqueda de valores atípicos, se busca estadísticas entre los datos, se normalizan los datos si hiciera falta, se seleccionan las instancias de data (observaciones) más útiles para el proyecto, se seleccionan las variables (o dimensiones) más útiles para el proyecto, se limpian los datos de ruidos o outliers que puedan entorpecer el estudio, entre otros procesos.
- **Modelado:** se seleccionan y aplican varios algoritmos de minería de datos y se

calibran los parámetros para obtener óptimos resultados. Hay varias técnicas que tienen requerimientos específicos para la forma de los datos, por lo que frecuentemente es necesario volver a la fase de preparación de datos. De esta fase se suele obtener modelos, que deben ser evaluados (indicar que tan bueno/preciso es el modelo)

- **Evaluación:** de los resultados del estudio se estudian los resultados desde una perspectiva de análisis de datos y se determina si los resultados satisfacen los requisitos, si el proceso fue llevado a cabo correctamente y se determinan cuáles eran los próximos pasos.
- **Despliegue:** Se dan a conocer los resultados, presentándolos en un formato legible y útil para la parte(s) interesada(s), puede ir acompañado de un plan de implementación, monitoreo y mantenimiento. En esta fase se entrega el informe final del proyecto y se valora al proyecto.

3.1.2 Métodos Ágiles

Son un grupo de métodos para el desarrollo de software que siguen los valores y principios del Manifiesto Ágil (ver figura 14). Los requerimientos y las soluciones evolucionan a lo largo del desarrollo. Estos métodos promueven el trabajo en equipo, colaboración, planificación adaptativa, desarrollo evolutivo, entrega temprana, mejora continua y fomenta la respuesta rápida y flexible a cambios en el desarrollo del software a lo largo del ciclo de vida del proyecto.

Los métodos ágiles poseen las siguientes características [20]:

- **Iterativo, incremental y evolutivo:** La mayoría de los métodos ágiles dividen las tareas en pequeñas partes con una planificación mínima y no involucran directamente una planificación a largo plazo. Las iteraciones incluyen un equipo multidisciplinario que trabaja en todas las funciones: planificación, análisis de requerimientos, diseño, programación, pruebas unitarias y pruebas de validación. Al final de cada iteración el producto es demostrado al cliente. Esto minimiza el riesgo general y permite al proyecto adaptarse a cambios rápidamente. Una iteración puede que no añada suficiente funcionalidad para garantizar un producto listo al mercado pero el objetivo es tener una versión disponible al final de cada iteración. Varias iteraciones son requeridas para tener un producto completamente listo o añadir nuevas funcionalidades al mismo.

- **Eficiente y comunicación cara-a-cara:** Cada equipo ágil debe tener comunicación directa con el cliente o al menos un representante del cliente. Esta persona es designada por el cliente para actuar en su nombre y dar su compromiso de estar disponible a los desarrolladores para preguntas durante las iteraciones. Al final de cada iteración, el cliente (o el representante) revisan el progreso y reevalúan las prioridades con el objetivo de optimizar el retorno de la inversión y asegurar el alineamiento con las necesidades del cliente y metas de la compañía. Ciclo de retroalimentación y adaptación muy breve: Una característica común del desarrollo ágil son las reuniones diarias de estado. Estas reuniones son cortas y usualmente antes de que el equipo comience a trabajar en el proyecto. Los miembros del equipo reportan entre ellos que hicieron el día anterior, que planean hacer a continuación y cuáles son sus obstáculos actualmente.
- **Enfoque de calidad:** Técnicas y herramientas específicas, así como integración continua, pruebas unitarias automatizadas, programación en pareja, desarrollo guiado por pruebas, patrón de diseño, diseño guiado por las necesidades del negocio y refactorización. La refactorización es el proceso de cambiar el diseño del código sin cambiar su comportamiento, se cambia cómo funciona el código sin alterar su resultado. Un punto importante de la refactorización es que ayuda a entender qué se debe cambiar. Al refactorizar, se procede en series de pequeñas transformaciones. Para conseguir algo significativo, se deben juntar varias refactorizaciones. Cada transformación es controlada y es de poco tiempo su realización. Esta cualidad es crucial en los métodos ágiles para manejar la flexibilidad del desarrollo de software. Los métodos ágiles se enfocan en diferentes aspectos del ciclo de vida para el desarrollo del software. Algunos se enfocan en administrar el desarrollo, como es el caso de Scrum y otros en implementar diferentes prácticas. Este último enfoque lo trata un método ágil llamado Extreme Programming o XP, el cual será profundizado en esta investigación.

Programación Extrema (XP)

Programación Extrema (Extreme Programming o XP) es un método ágil de desarrollo de software enfocado en la satisfacción del usuario, planeado para mejorar la calidad y capacidad de respuesta a los cambios de requerimientos del usuario [21]. Fue creado por Kent Beck en Octubre de 1999 cuando publicó Extreme Programming Explained. Es una

forma ligera, eficiente, flexible, predecible, científica y de bajo riesgo para desarrollar software. Siendo un método ágil, propone las entregas frecuentes en ciclos de desarrollo cortos, que tienen como propósito mejorar la productividad e introducir puntos de control en donde se pueden adoptar nuevos requerimientos del usuario.

Se distingue de otras metodologías por:

- Su retroalimentación temprana, concreta y continúa de sus ciclos cortos.
- Su enfoque de planificación incremental, que rápidamente crea un plan general que se espera que evolucione a lo largo de la vida del proyecto.
- Su habilidad de ser flexible con la implementación de funcionalidades, respondiendo a los cambios en las necesidades del negocio.
- Su dependencia en pruebas automatizadas hechas por programadores y usuarios para monitorear el progreso del desarrollo, permitiendo al sistema evolucionar y detectar errores de manera temprana.
- Su dependencia en la comunicación oral, pruebas y código fuente para comunicar la intención y estructura del sistema.
- Su dependencia en su proceso de diseño evolutivo que persiste tanto como el sistema necesite.
- Su dependencia en una colaboración muy unida de programadores con habilidades ordinarias.
- Su dependencia en prácticas que trabajan tanto con los instintos a corto plazo de los programadores y los intereses a largo plazo del proyecto.
- Extreme Programming es también una disciplina de desarrollo de software que sigue una estructura específica que está diseñada para simplificar y acelerar el proceso para desarrollar nuevo software.

Según Beck, este método se guía por doce prácticas que son descritas a continuación.

1. El Proceso de Planificación: determinar el alcance de la entrega combinando prioridades del negocio y estimando tiempo de desarrollo. Si la realidad supera el plan, se actualiza el plan.

2. Pequeñas entregas: poner un sistema simple en producción rápidamente, luego realizar nuevas versiones en un corto periodo de tiempo.
3. Metáfora: guiar todo el desarrollo compartiendo una historia simple de como el sistema funciona.
4. Diseño Simple: el sistema debe ser diseñado lo más simple posible. Complejidad extra es removida tan pronto como es encontrada.
5. Pruebas: los programadores escriben pruebas que deben correr perfectamente para que el desarrollo continúe.
6. Refactorización: los programadores mejoran el diseño del software durante cada etapa del progreso en vez de esperar hasta el final y volver para corregir errores.
7. Programación en Parejas: todo el código está escrito por una pareja de programadores trabajando en el mismo computador. Dos personas trabajando juntas en un computador entregan trabajos de mejor calidad que al trabajar separados. Se sientan lado a lado al frente del monitor y comparten tanto el teclado como el ratón. Concentrados en el código que se está escribiendo.
8. Propiedad Colectiva: cada línea de código pertenece a cada programador trabajando en el proyecto, de esta forma no existen problemas de autoría propia que atrasen el desarrollo. El código se cambia cuando se necesita cambiar, sin retraso.
9. Interacción Continua: el equipo integra y construye el sistema de software varias veces al día para mantener a todos los programadores en la misma etapa del proceso de desarrollo.
10. 40-Horas Semanales: el equipo no trabaja excesivamente para asegurar que se mantiene bien descansado, alerta y efectivo. Si se trabaja excesivamente una semana, nunca repetir la semana próxima.
11. Cliente En-Lugar: el proyecto es dirigido por el cliente (o un representante del cliente) que está disponible todo el tiempo para responder preguntas, establecer prioridades y determinar los requerimientos del proyecto.
12. Estándar de Codificación: todos los programadores escriben el código de la misma manera. Esto les permite trabajar en pares y compartir autoría del código.

La forma tradicional de desarrollar código es lineal, con cada etapa del ciclo de desarrollo requiriendo terminación de la etapa anterior. Por lo que la Programación Extrema cambia este paradigma. El desarrollo comienza con la planificación y continua con el diseño, la codificación, las pruebas y como fase extra tenemos la retroalimentación.

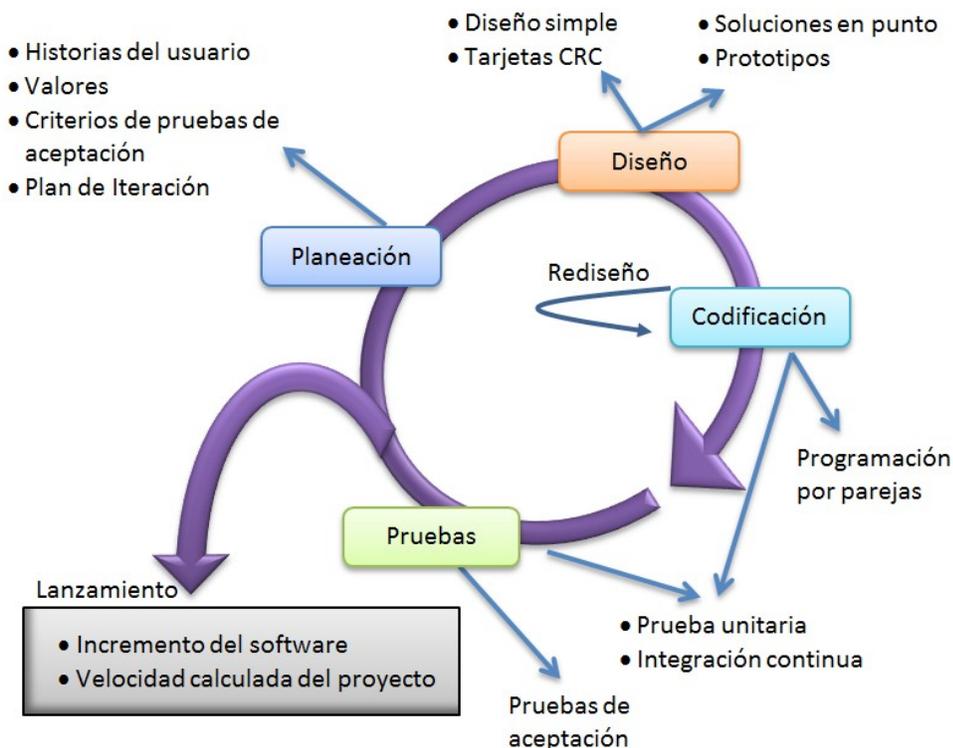


Figura10. Ciclo de desarrollo de Programación Extrema (XP)

A continuación se explican las etapas de este ciclo de desarrollo:

- **Planeación:** la primera fase donde los usuarios o clientes se reúnen con el equipo de desarrollo para crear los requerimientos. El equipo convierte estos requerimientos en iteraciones que cubren una pequeña parte de la funcionalidad o características necesarias. Una combinación de iteraciones provee al cliente con un producto final completamente funcional. El equipo de programación prepara el plan, tiempo y costos de llevar a cabo las iteraciones, desarrolladores individuales se registran para las iteraciones. Una planificación común es el Método de la Ruta Crítica, se agrupan iteraciones esenciales para el progreso del proyecto en una manera lineal y se organizan para terminación otras iteraciones paralelas a la ruta crítica.

- **Diseño:** los principios que guían esta etapa son:
 - Impulsar la simplicidad al manifestar un objeto solamente una vez y no añadir funcionalidad con anterioridad.
 - Utilizar el sistema de metáforas o estándares en nombres, clases y métodos, y estar de acuerdo en estilos uniformes y formatos para asegurar la compatibilidad entre el trabajo de diferentes miembros del equipo.
 - Creando programas simples que exploran las soluciones potenciales para un problema en específico, ignorando las demás preocupaciones, para mitigar riesgos.
- **Codificación:** constituye la fase más importante del ciclo de desarrollo. La programación extrema da prioridad a la codificación real sobre las otras tareas tal como la documentación para asegurar que el cliente recibe algo considerable en valor al final del día. Algunos de los estándares relacionados con la codificación incluyen:
 - Desarrollar el código basado en las metáforas y estándares acordados, y adoptar una política de autoría colectiva del código.
 - Programación en pareja, teniendo como objetivo producir código de alta calidad al mismo o menor costo.
 - Estricta fidelidad a las 40 horas semanales de trabajo sin sobretiempo. Esto asegura que los desarrolladores trabajen con sus habilidades mentales y físicas al máximo.
 - Integración frecuente del código al repositorio dedicado, solo con una pareja integrando al tiempo para prevenir conflictos y optimización al final.
- **Pruebas:** la Programación Extrema integra las pruebas en la fase de desarrollo a diferencia de al terminar la fase de desarrollo. Todo el código tiene unidades de pruebas para eliminar errores y el código debe pasar todas estas unidades de pruebas antes del lanzamiento. Otra prueba importante es las pruebas de aceptación del cliente, basadas en las especificaciones del cliente. Estas pruebas de aceptación son ejecutadas al terminar de codificar y los desarrolladores le entregan los resultados al cliente junto con las demostraciones.
- **Retroalimentación:** siendo una etapa extra como la base de la Programación Extrema es

el mecanismo de involucrar al cliente continuamente a través de la retroalimentación durante la fase de desarrollo. El desarrollador también recibe retroalimentación del gerente de proyecto. La base de esta etapa extra son las pruebas de aceptación del cliente. La Programación Extrema contiene una importante filosofía que trata con concentrarse en la calidad y el alcance. Concentrarse en la calidad puede agilizar el desarrollo cuando se construye software más fiable.

Crear una cantidad considerable de pruebas provee al equipo de una confianza para escribir código más rápido y con menos estrés al saber que no se va a estropear algo. Trabajar en un buen sistema anima al equipo. El otro enfoque está en el alcance. Al escribir software, preocuparse menos es una gran ayuda. Darse cuenta de los requerimientos mínimos viables crea mejor software sin atrasos. El cliente no necesariamente sabe que quiere al comienzo, al ver software real pueden refinar y limitar su alcance de los requerimientos.

3.2. HERRAMIENTAS A UTILIZAR

A continuación se presentan las herramientas utilizadas para construir la lógica de negocio en una aplicación. Esto implica el procesamiento de las solicitudes que se realizan desde la interfaz de usuario a los controladores así como la comunicación con el módulo encargado del manejo de datos.

- **Lenguajes**
 - ✓ R
 - ✓ PHP
 - ✓ HTML
 - ✓ CSS
 - ✓ JAVASRIPT
 - ✓ JQUERY

- **IDE**
 - ✓ R-Studio
 - ✓ NetBeans

- **Paquetes de R**

Conjunto de instrucciones y funciones, las cuales permiten que los programas en R realicen distintas operaciones. Entre los paquetes que se están utilizando se encuentran:

- o TwitterR: Proporciona una interfaz para el API Web Twitter.
- o RMongo: Interfaz de base de datos MongoDB para R. Se proporciona la interfaz a través de Java llama a la mongojava conductor.
- o Stringr: Consiste en etiquetas para el manejo sencillo de operaciones entre strings.
- o Tm: Un marco de trabajo para aplicaciones de minería de texto dentro de R.
- o SnowballC: Una interfaz para la librería C libstemmer que implementa derivación de palabras hasta su raíz común para comparar vocabulario, actualmente soporta los lenguajes danés, holandés , Inglés , finlandés, francés, alemán, húngaro, italiano, noruego , portugués, rumano, ruso , español , sueco y turco.
- o Wordcloud: Genera nubes de palabras.
- o Fpc: Procedimientos flexibles para Clustering.
- o Igraph: Análisis y visualización de grafos
- o NLP: Infraestructura para procesamiento del Lenguaje Natural.
- o RcolorBrewer: Provee esquema de colores para mapas y gráficos.
- o Rjava: Interfaz de bajo nivel para la máquina virtual de java, permite la creación de objetos, llamar a los métodos y acceso a los campos.

- **Librerías de PHP**

- o php5-mongo

- **Base de Datos no relacionales**

- o MongoDB

- **API**

- o RESTful API

CAPITULO 4

DESARROLLO DE LA SOLUCIÓN

4.1. ARQUITECTURA DE LA SOLUCIÓN

El proyecto se dividió en tres partes fundamentales, un proceso de ETL donde se extrae, se transforma y se almacena la información extraída desde Twitter en una base de datos MongoDB, la siguiente parte es un proceso de minería y por último el sistema web.

Primera parte: Proceso ETL:

1. Se creó una aplicación de Twitter, con esta aplicación se tiene acceso a la API de Twitter, esto es necesario para poder extraer la información del microblogging.
2. Una vez creada la aplicación se procede a utilizar las claves que provee dicha aplicación con el lenguaje de programación R, para esto se utiliza el IDE Rstudio y particularmente el paquete `twitteR`, este paquete contiene las funciones necesarias para acceder a la información contenida en Twitter utilizando R.
3. Se aplican técnicas de limpieza y transformación de los tweets, usando las funciones de los paquetes que tiene el lenguaje R (como son los paquetes `tm` y `stringr`), se eliminan los números, los símbolos, los emoticones y los enlaces a otras páginas porque no es información relevante para el proceso de análisis.
4. Por último se guarda la información limpia en formato JSON en la base de datos MongoDB, para ello se utiliza el paquete `Rmongo`, además se configura la base de datos para tener una mayor seguridad, utilizando roles, usuarios, permisos de usuarios y contraseñas.

Segunda parte: proceso de minería:

1. Se utilizó el lenguaje R, ya que contiene todas las funcionalidades de limpieza de

datos, transformación, agrupación y diversas formas de mostrar resultados gráficamente que muestran la información relevante contenida en los tweets.

2. Consultamos la información contenida en MongoDB y la transformamos en un dataframe, el cual es, una tabla que contiene en cada columna los nombres de los atributos y en cada fila cada registro.
3. Se eliminan los caracteres extraños, signos de puntuación y se lleva a minúsculas todas las palabras.
4. Si el caso es generar una nube de palabras de los tweets seleccionados entonces:
 - 4.1 Eliminamos las palabras que no aporten información relevante al análisis, como son los artículos, pronombres y conectores.
 - 4.2 Creamos una estructura de matriz términos-documentos, esta estructura permite conocer cuántas veces aparece una palabra (término) en cada tweet (documento).
 - 4.3 Almacenamos en un dataframe las palabras y sus frecuencias, con esto se puede utilizar el paquete wordcloud para generar la imagen de la nube de palabras.
5. Si el caso es generar un histograma de polaridad:
 - 5.1 Consultamos la base de datos para obtener la información de las palabras positivas y negativas seleccionadas.
 - 5.2 Eliminamos las palabras que no aporten información relevante al análisis, como son los artículos, pronombres y conectores.

5.3 Buscamos cada palabra positiva y negativa dentro de cada tweet, si coincide con alguna palabra positiva se suma uno (+1) a la polaridad de ese tweet, si coincide con alguna palabra negativa entonces se resta uno (-1) a su polaridad, de esta forma se calculan las polaridades.

5.4 Creamos un dataframe con los tweets y sus polaridades, con esto se puede crear un histograma personalizado donde se muestra gráficamente la información, esto se hace con el paquete ggplot2.

Tercera parte: proceso de desarrollo del sistema web:

Este proceso integra la parte ETL, minería, visualización de la información y funcionalidades administrativas del sistema. Se utilizó el lenguaje PHP con el IDE Netbeans, bajo el modelo de desarrollo web MVC, modelo-vista-controlador.

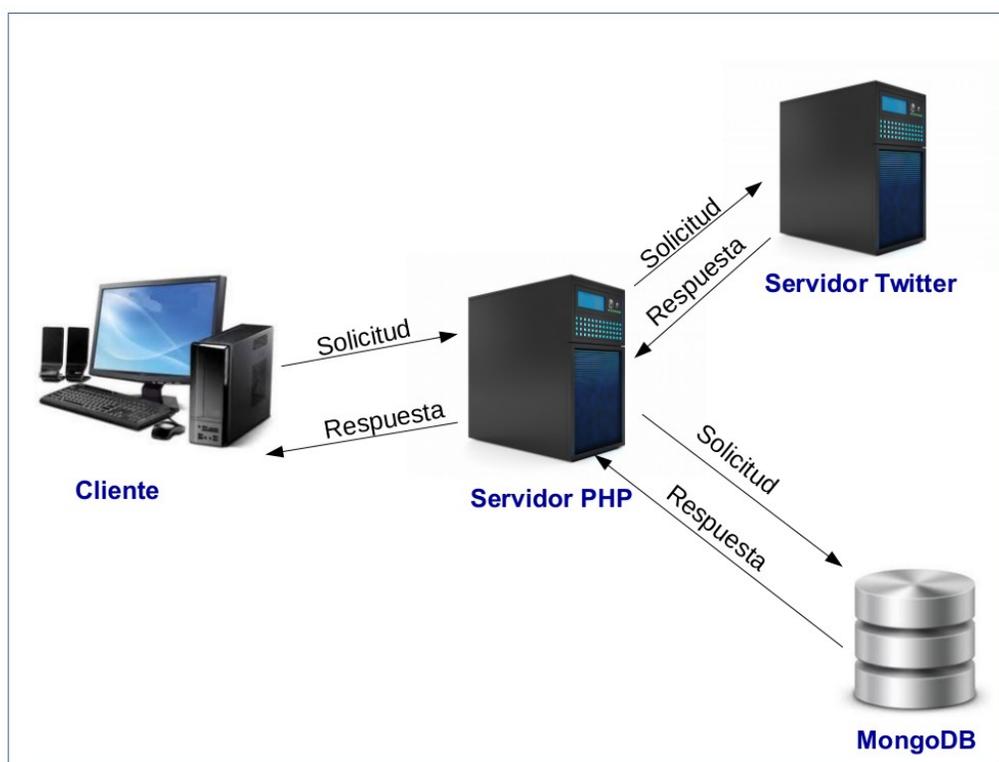


Figura 11. Arquitectura de la solución

Las características técnicas de la laptop utilizada son: Debian 8 Jessie, i5 de 4

núcleos, 4GB de RAM, 500 GB de disco duro.

Resumiendo la arquitectura de la solución está formada por tres grandes procesos, todos creados bajo software libre R, PHP y MongoDB. El modelo de desarrollo es MVC y los IDE's utilizados son Rstudio y Netbeans.

4.2. ANÁLISIS Y DISEÑO DE LA SOLUCIÓN

Una vez definidas las metodologías que utilizaremos nos guiamos de ellas para llevar a cabo la aplicación y creamos un calendario de actividades el cual nos permitió que el usuario final fuera viendo los avances que se llevaban por día.

N°	Objetivo	Actividad	Fecha Inicio	Fecha Fin
1	Levantamiento de requerimientos	Levantamiento de requerimientos	28/03/2016	04/04/2016
2	Modelado de la solución	Modelo de la solución	04/04/2016	11/04/2016
3	Selección de las herramientas apropiadas para el desarrollo del sistema	Análisis de las herramientas a utilizar para el desarrollo del sistema	11/04/2016	18/04/2016
4	Instalación y configuración de las herramientas seleccionadas	Instalación y configuración de las herramientas para el desarrollo del sistema	18/04/2016	25/04/2016
5	Crear conexión entre R y Twitter	Crear aplicación de twitter, crear conexión entre Twitter y R utilizando paquete twitterR	25/04/2016	02/05/2016
6	Limpieza de la información extraída de Twitter	Descargar tweets dada una cantidad y guardarlos en un dataframe selección de columnas para	02/05/2016	09/05/2016

		el análisis y eliminar caracteres extraños y enlaces del texto de cada tweet		
7	Crear conexión entre R y MongoDB	Crear conexión entre R y MongoDB, almacenar los tweets en formato JSON en MongoDB usando la librería Rmongo	09/05/2016	16/05/2016
8	Limpieza de la información extraída de Twitter, crear conexión entre R y MongoDB	Consultar la base de datos, obtener los tweets almacenados y guardarlos en un dataframe, realizar limpieza de texto de los tweets (letras acentuadas por letras sin acento, llevar todo a minúsculas, eliminar los números, etc), utilizar archivo de limpieza (stopwords) para eliminar las palabras que no dan información relevante al análisis	16/05/2016	23/05/2016
9	Crear nube de palabras utilizando la	Ordenar palabras dada su cantidad de	23/05/2016	30/05/2016

	información ya procesada		menciones de forma decreciente, crear nube de palabras y guardar la imagen en una carpeta predeterminada		
10	Crear histograma de polaridad utilizando la información ya procesada		Consulta de colección de palabras positivas y negativas, utilizar palabras negativas y calcular polaridad de cada tweet, hacer histograma de polaridad de los tweets encontrados	30/05/2016	06/06/2016
11	Crear histograma de polaridad utilizando la información ya procesada		Personalizar histograma de polaridad utilizando la librería de ggplot2 de R	06/06/2016	13/06/2016
12	Conectar PHP y R		Crear Rscripts que se ejecuten directamente por comandos de consola, ajustar las funciones de nube de palabras e histograma de polaridad en forma de Rscripts	13/06/2016	20/06/2016
13	Conectar PHP y R		Crear formato para recibir parámetros de	20/06/2016	27/06/2016

			los Rscripts, modificar contenido de la carpeta que contiene los paquetes por defecto de R para que los Rscripts utilicen todos los paquetes de R		
14	Conectar PHP y R	PHP	Realizar pruebas de las funciones de nube de palabras e histograma de polaridad en forma de Rscripts recibiendo parámetros	27/06/2016	04/07/2016
15	Crear conexión entre PHP y MongoDB	PHP	Crear conexión ente PHP y MongoDB	04/07/2016	11/07/2016
16	Utilizar modelo de desarrollo MVC para crear el sistema	el de	Crear vista, Javascript y controlador de inicio de sesión	11/07/2016	18/07/2016
17	Utilizar modelo de desarrollo MVC para crear el sistema	el de	Crear vista principal para los usuarios, crear vista de información del sistema	18/07/2016	25/07/2016
18	Utilizar modelo de desarrollo MVC para crear el sistema	el de	Crear vista, Javascript y controlador de cambio de contraseña junto con envió de correo de la nueva clave	25/07/2016	01/08/2016
19	Utilizar modelo de	el de	Crear vista, Javascript y	01/08/2016	08/08/2016

	desarrollo MVC para crear el sistema	controlador de cambio de cargar búsqueda, sea por hoja de cálculo o de forma manual		
20	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para gestionar las búsquedas, modificando el cron del equipo	08/08/2016	15/08/2016
21	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para modificar archivos de limpieza.	15/08/2016	22/08/2016
22	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para modificar palabras positivas, crear vistas, Javascripts y controladores para modificar palabras negativas	22/08/2016	29/08/2016
23	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para ver resultados	29/08/2016	05/09/2016
24	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para modificar archivos de limpieza por defecto	05/09/2016	12/09/2016
25	Utilizar el modelo de desarrollo	Crear vistas, Javascripts y controladores	12/09/2016	19/09/2016

	MVC para crear el sistema	para modificar palabras positivas por defecto, crear vistas, Javascripts y controladores para modificar palabras negativas por defecto		
26	Utilizar el modelo de desarrollo MVC para crear el sistema	Crear vistas, Javascripts y controladores para modificar usuarios	19/09/2016	26/09/2016
27	Utilizar el modelo de desarrollo MVC para crear el sistema	Realizar pruebas sobre las funciones del usuario administrador, realizar pruebas sobre las funciones del usuario analista y modificar vistas de resultados para mayor usabilidad del usuario	03/10/2016	10/10/2016

En el presente trabajo especial de grado se diseñó el siguiente modelo de base de Datos.

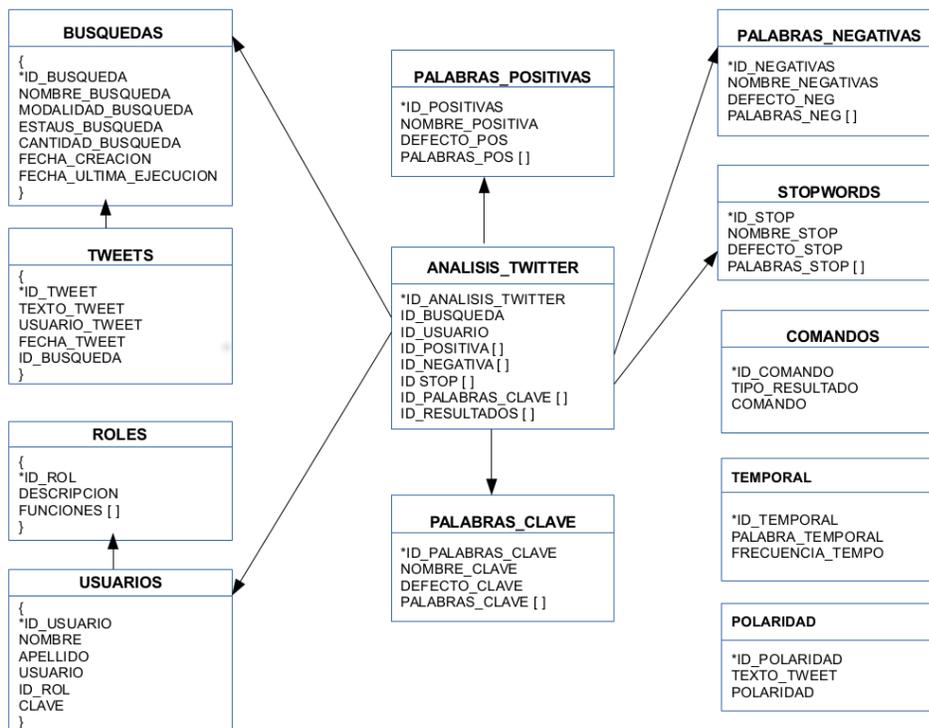


Figura 12. Modelo de Base de Datos.

Se creó un diagrama de casos de uso que nos muestra las funcionalidades con las cuales el sistema contara tanto para el usuario analista como para el usuario administrador. En cual podrán observar en la siguiente figura.

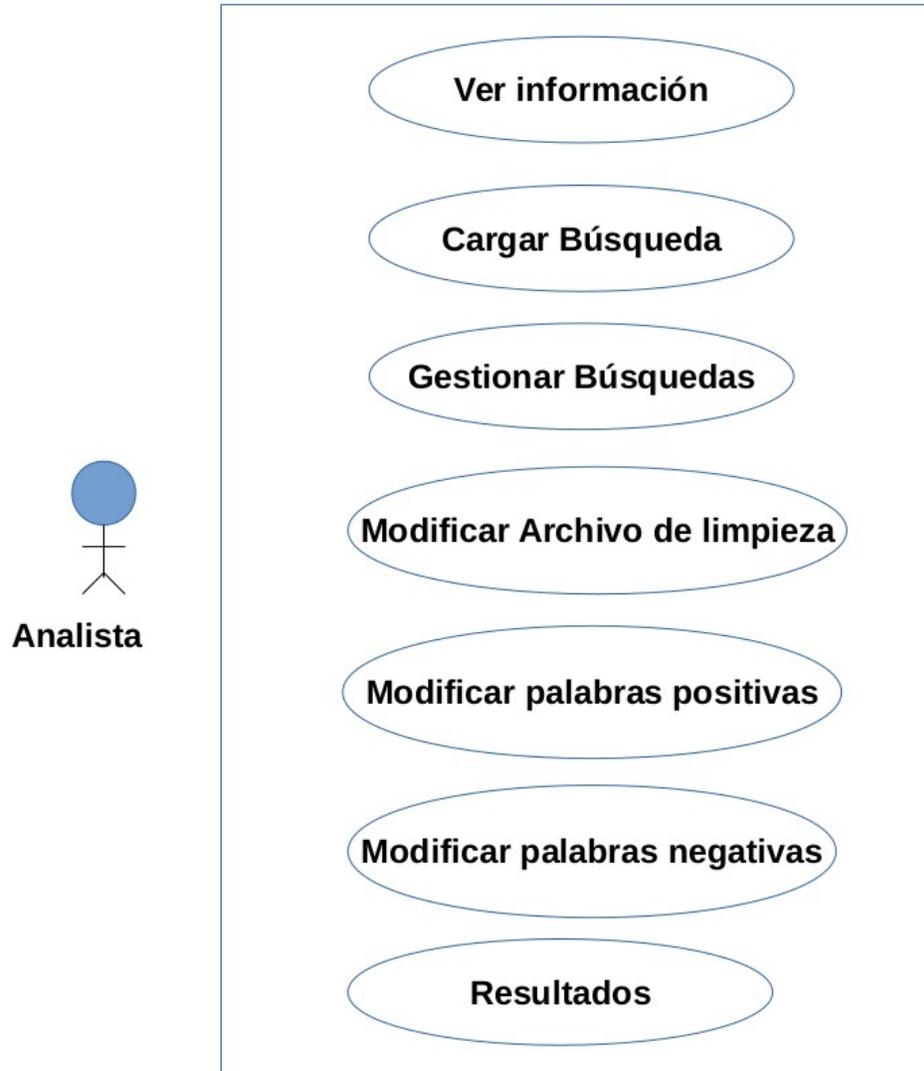


Figura 13. Diagrama de casos de uso Analista

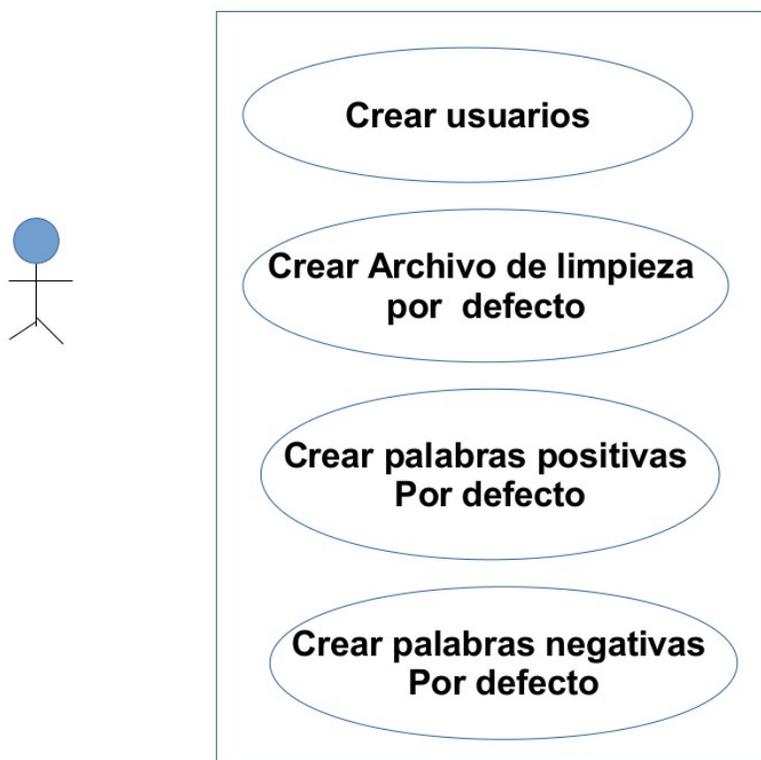


Figura 13.1. Diagrama de casos de uso Administrador

Se crearon diseños para las vista de la aplicación las cuales podremos observar en las siguientes figuras. A continuación mostraremos las vistas del usuario analista.



Figura 14. Principal.

El sistema cuenta una interfaz para cargar búsquedas de dos maneras la primera es cargar búsquedas por archivo, la cual permite seleccionar un archivo que contienen un grupo de palabras a buscar, es decir cantidades de tweets que contengan las palabras específicas en el archivo, este archivo es una hoja de cálculo. Que debe cumplir el siguiente formato:

- El nombre del archivo siempre debe comenzar con “archivo_busqueda_”
- Y dentro el archivo debe contener las siguientes columnas: Palabras , Cantidad.

A	B	C	D
Palabras	Cantidad		
chacao	50		
hueco	50		
tronera	50		
calle	50		

Figura 15. Formato del archivo de búsquedas.

Y la segunda forma de cargar las búsquedas es de forma manual. En la cual debe ingresar la palabra a buscar, la cantidad y el nombre para la búsqueda. En la siguiente figura podrán observar la interfaz correspondiente a la búsqueda.

Cargar Búsqueda

Cargar Búsqueda Por Archivo

Seleccionar archivo No se eligió

Cargar Búsqueda

Cargar Búsqueda de Forma Manual

Palabra:

Cantidad:

Nombre de la Búsqueda:

Cargar Búsqueda

Figura 16. Cargar búsquedas

Se cuenta con una interfaz para gestionar las búsquedas la cual permite iniciar una búsqueda, seleccionar el rango de ejecución de la búsqueda ya sea una sola vez, por horas, días, semana o mes. Y a la vez se puede pausar, reanudar y culminar la búsqueda. En la siguiente figura podrán ver la interfaz de gestionar búsquedas.

Gestionar Búsquedas



The screenshot shows a web interface for managing searches. At the top left, there is a blue button labeled 'Gestionar Búsquedas'. Below it is a table with a dark blue header titled 'BÚSQUEDAS'. The table has six columns: 'Nombre de la Búsqueda', 'Estatus de la Búsqueda', 'Modalidad de la Búsqueda', 'Acciones', 'Fecha de Creación de la Búsqueda', and 'Fecha Ultima Ejecución de la Búsqueda'. There are four rows of data representing different searches.

Nombre de la Búsqueda	Estatus de la Búsqueda	Modalidad de la Búsqueda	Acciones	Fecha de Creación de la Búsqueda	Fecha Ultima Ejecución de la Búsqueda
hueco	SIN INICIAR		Iniciar	2016-10-03 13:39:08	
chacao	EN PROCESO	HORA	Pausar Terminar	2016-10-03 13:39:20	
tronera	EN PAUSA	HORA	Reanudar Terminar	2016-10-03 13:39:51	
calle	CULMINADA	UNICA		2016-10-03 13:40:01	2016-10-03 13:41:23

Figura 17. Gestionar búsquedas.

El sistema permite agregar, modificar, ver y eliminar los Archivos de limpieza. Vale destacar que solamente se podrán eliminar y modificar los archivos que no son por defecto.

El formato de los archivos de limpieza debe cumplir con las siguientes especificaciones, la primera línea debe decir Palabras y el nombre con el cual será guardado es `archivo_limpieza_`.

En la siguiente figura podrán observar un ejemplo de un archivo de limpieza

A	B	C	D
Palabras			
a			
al			
algo			
algunas			
algunos			
ante			
antes			
como			
con			
contra			
cual			
cuando			
de			
del			
desde			
donde			
durante			
e			
el			
ella			
ellas			
ellos			
en			
entre			
era			
erais			
éramos			
eran			
eras			
eres			
es			
esa			
esas			
ese			
eso			
esos			
esta			
está			
estaba			
estabais			
estabamos			
estaban			
estabas			
estad			
estada			

Figura 18. Formato del Archivo de limpieza.

A continuación podrán ver la interfaz de agregar, modificar y eliminar archivo de limpieza.

Modificar Stopwords

No se eligió archivo

STOPWORDS			
Nombre del Stopwords			
stopwords por defecto	<input type="button" value="Mostrar Detalle Stopwords"/>		
español	<input type="button" value="Mostrar Detalle Stopwords"/>		
ensayo	<input type="button" value="Mostrar Detalle Stopwords"/>	<input type="button" value="Modificar Stopwords"/>	<input type="button" value="Eliminar Stopwords"/>

Figura 19. Mostrar archivo de limpieza.

DEFECTO	
Nº	Palabra
11	accesible
12	acelerado
13	aclamación
14	aclamado
15	aclarar
16	acogedor
17	acomodativa
18	acreditado
19	actualizable
20	actualizado

Mostrando 11 - 20 de 1995 Mostrar por página: 10 | 20 | 40 Pagina

Figura 20. Detalle Palabras positivas.

El sistema permite modificar palabras negativas, funciona de la misma forma que la interfaz de modificar palabras positivas y modificar archivos de limpieza.

Modificar Palabras Negativas

Mostrar Palabras Negativas
Seleccionar archivo | No se eligió archivo
Agregar Palabras Negativas

PALABRAS NEGATIVAS			
Nombre de las Palabras Negativas			
defecto	Mostrar Detalle Negativas		

Figura 21. Mostrar Palabras negativas

DEFECTO	
Nº	Palabra
1	a cuadros
2	a distancia
3	a escondidas
4	a regañadientes
5	a traición
6	abadejo
7	abandonado
8	abandonar
9	abandonos
10	abaratar

Mostrando 1 - 10 de 4705 Mostrar por página: 10 | 20 | 40 Pagina 1 2 3 .. 471

Figura 21.1. Detalle Palabras negativas

El sistema cuenta con una interfaz que genera resultado tanto por nubes de palabras como por histograma de polaridad que se mostrar en la siguiente figura.

Emmanuel Galeano

Resultados

Resultados por Búsquedas

- Nube de Palabras
- Histograma de Polaridad
- Texto de los Tweets

BUSQUEDAS:

rivas_davila
 rivas_davila_2
 rivas_davila_3
 rivas_davila_4
 rivas_davila_5
 puente_rastrojos
 puente_rastrojos2

ARCHIVOS DE LIMPIEZA:

stopwords por defecto
 español
 ensayo

Generar nube

Figura 22. Resultados por Nube de Palabras



Figura 23. Resultados por Histograma de polaridad

Continuamos con las interfaces del usuario administrador, que a diferencia del usuario analista este podrá:

- Modificar, crear y eliminar archivos de limpiezas y archivo de palabras positivas y negativas predefinidas por el sistema.
- Crear usuarios.

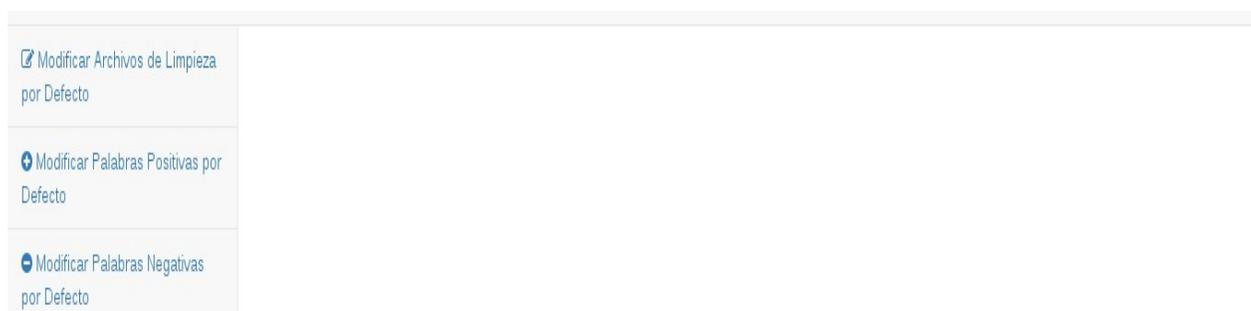


Figura 24. Vista Principal Administrador.

[Ver Archivos de Limpieza](#)

 No se eligió archivo

ARCHIVOS DE LIMPIEZA			
Nombre del Archivo de Limpieza			
stopwords por defecto	<input type="button" value="Mostrar Detalles"/>	<input type="button" value="Modificar"/>	<input type="button" value="Eliminar"/>
español	<input type="button" value="Mostrar Detalles"/>	<input type="button" value="Modificar"/>	<input type="button" value="Eliminar"/>
ensayo	<input type="button" value="Mostrar Detalle"/>		
ensayo_defecto	<input type="button" value="Mostrar Detalles"/>	<input type="button" value="Modificar"/>	<input type="button" value="Eliminar"/>

Figura 25 Modificar archivo de limpieza Administrador.

Modificar Palabras Positivas por Defecto

[Ver Palabras Positivas](#)

 No se eligió archivo

PALABRAS POSITIVAS			
Nombre de las Palabras Positivas			
ensayo_defecto	<input type="button" value="Mostrar Detalle Positivas"/>	<input type="button" value="Modificar Positivas"/>	<input type="button" value="Eliminar Positivas"/>
defecto	<input type="button" value="Mostrar Detalle Positivas"/>		

Figura 26 Modificar Palabras por defecto Administrador.

Modificar Palabras Negativas por Defecto

PALABRAS NEGATIVAS			
Nombre de las Palabras Negativas			
defecto	Mostrar Detalle Negativas		
ensayo_defecto	Mostrar Detalle Negativas	Modificar Negativas	Eliminar Negativas

Figura 27 Modificar Palabras Negativas por Defecto Administrador.

Nombre del Usuario	Apellido del Usuario	Correo del Usuario	Rol del Usuario		
Emmanuel	Galeano	egaleano@cfg.gob.ve	Analista	Modificar	Eliminar
Irima	Rodriguez	irodriguez@cfg.gob.ve	Administrador		
edward	vieras	evieras@cfg.gob.ve	Administrador	Modificar	Eliminar

Figura 28 Agregar, Modificar y Eliminar usuarios.

4.3 DESARROLLO

4.3.1. ANÁLISIS EXPLORATORIO DE LOS DATOS

Teniendo toda la metadata era necesario extraer el texto de los tweets y llevarlos a una estructura tipo tabla llamada dataframe, que pudiera ser utilizada por el programa en R. El dataframe es una tabla que cumple las mismas propiedades de los csv (coma separate values). Para eso se utilizó la función “twListstToDF” que transforma un conjunto de tweets en un dataframe, a este dataframe se le extrajo los atributos, nombre de usuario, fecha de publicación y texto asociado a cada tweet, para así crear una cadena de caracteres en formato JSON por cada tweet, esta cadena sera la que se almacene en la base de datos mongoDB y así poder consultar la información.

	id_búsqueda	fecha_tweet	texto_tweet	X_id	scri
224	5	2016-10-01 22:04:19	RT Merida171 PRECAUCIÓN Nos reportan fuertes p...	57f6badf44aee70709e93021	EIMe
225	5	2016-10-06 14:54:34	Para el 20 MP Rivas Dávila del EstadoMérida tendrá...	57f6bae244ae140a881db332	Surr
226	5	2016-10-03 22:30:10	RT INMIVI GobAlexisR construcción puente los rast...	57f6bae244ae140a881db333	com
227	5	2016-09-28 20:20:45	Siembra de plántulas de lechuga y ají dulce UE Ge...	57f6bae244ae140a881db334	ptm
228	5	2016-09-28 18:25:26	RT INMIVI GobAlexisR construcción puente los rast...	57f6bae244ae140a881db335	Chlc
229	5	2016-09-28 04:38:07	GobdeMerida FONHVM RT INMIVI GobAlexisR const...	57f6bae244ae140a881db336	rdfa
230	5	2016-09-28 04:26:04	FONHVM RT INMIVI GobAlexisR construcción puent...	57f6bae244ae140a881db337	Gob
231	5	2016-09-28 04:21:26	RT INMIVI GobAlexisR construcción puente los rast...	57f6bae244ae140a881db338	FON
232	5	2016-09-28 03:13:07	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db339	Moc
233	5	2016-09-28 01:36:03	RT INMIVI GobAlexisR construcción puente los rast...	57f6bae244ae140a881db33a	Inju
234	5	2016-09-28 01:02:05	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db33b	Fern
235	5	2016-09-27 22:25:19	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db33c	Jose
236	5	2016-09-27 22:15:24	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db33d	460i
237	5	2016-09-27 21:14:04	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db33e	evgt
238	5	2016-09-27 18:34:55	RT INMIVI GobAlexisR construcción puente los rast...	57f6bae244ae140a881db33f	ARA
239	5	2016-09-27 12:33:45	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db340	mis
240	5	2016-09-27 04:20:47	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db341	day
241	5	2016-09-27 03:54:42	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db342	Bloc
242	5	2016-09-27 03:02:03	RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE...	57f6bae244ae140a881db343	clari
243	5	2016-10-06 20:55:42	NOTIDIARIO007 Marquina Maduro aprobó Bs 15 bill...	57f6bae544ae0f417a4b5f8b	Beis
244	5	2016-10-06 20:55:38	RT ivanrive3 Merida 410pm AHORA Transportistas ...	57f6bae544ae0f417a4b5f8c	Mar
245	5	2016-10-06 20:45:47	RT AloCiudadano Marquina Maduro aprobó Bs 15 b...	57f6bae544ae0f417a4b5f8d	mar

Showing 223 to 245 of 342 entries

Figura 29. Dataframe

Del dataframe se extrajeron los atributos text, created y screenName. Ya que los otros atributos no aportaban información relevante para el análisis de texto ni para la realización de reportes puesto que eran valores verdaderos o falsos, números o valores faltantes como por ejemplo las retwiteado o no, el identificador del tweet y coordenadas geográficas.

- text: Cadena de caracteres que contiene el texto del tweet.
- created: Fecha de publicación del tweet.
- screenName: cadena de caracteres que indica el nombre del usuario que creó el tweet.

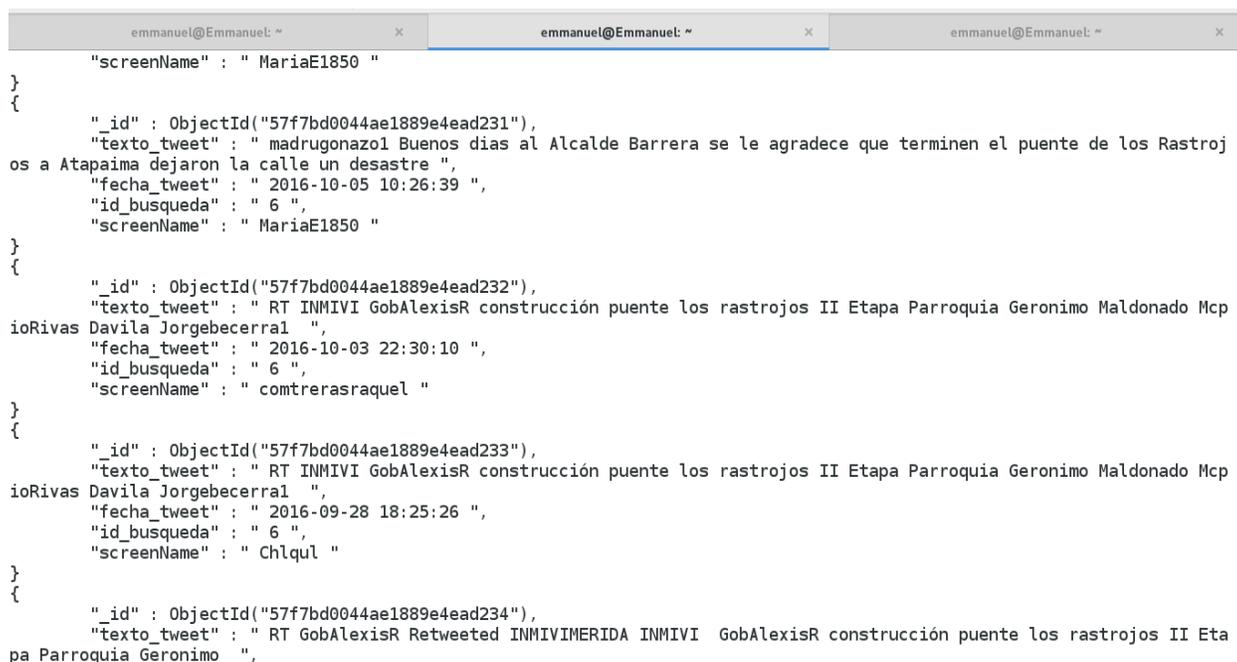
4.3.2. TRANSFORMACIÓN DE LA DATA

Para realizar el proceso de transformación de los datos en nuestro trabajo especial de grado lo que se hizo fue extraer los atributos text, created y screenName, convertirlos en una cadena de caracteres que cumpliera con el formato JSON. Al momento de insertar la

información en la base de Datos se agrega automáticamente un identificador a cada tweet, además se agregó el atributo `id_búsqueda` el cual se utiliza para asociar a que búsqueda pertenece el tweet.

Se renombraron los atributos para facilitar su uso en la aplicación, quedando cada documento con los siguientes nombres de atributos:

- `_id`
- `texto_tweet`
- `fecha_tweet`
- `id_búsqueda`
- `screenName`



```

"screenName" : " MariaE1850 "
}
[
  {
    "_id" : ObjectId("57f7bd0044ae1889e4ead231"),
    "texto_tweet" : " madrugonazo1 Buenos dias al Alcalde Barrera se le agradece que terminen el puente de los Rastrojos a Atapaima dejaron la calle un desastre ",
    "fecha_tweet" : " 2016-10-05 10:26:39 ",
    "id_búsqueda" : " 6 ",
    "screenName" : " MariaE1850 "
  }
  {
    "_id" : ObjectId("57f7bd0044ae1889e4ead232"),
    "texto_tweet" : " RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado Mpio Rivas Davila Jorgebecerral ",
    "fecha_tweet" : " 2016-10-03 22:30:10 ",
    "id_búsqueda" : " 6 ",
    "screenName" : " comtrerasraquel "
  }
  {
    "_id" : ObjectId("57f7bd0044ae1889e4ead233"),
    "texto_tweet" : " RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado Mpio Rivas Davila Jorgebecerral ",
    "fecha_tweet" : " 2016-09-28 18:25:26 ",
    "id_búsqueda" : " 6 ",
    "screenName" : " Chlqu "
  }
  {
    "_id" : ObjectId("57f7bd0044ae1889e4ead234"),
    "texto_tweet" : " RT GobAlexisR Retweeted INMIVIMERIDA INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo "
  }
]

```

Figura 30. Tweets en formato JSON

4.3.3. RESULTADOS

Para la visualización de resultados se pueden seleccionar por búsquedas, las cuales permiten trabajar con el universo completo de los tweets encontrados.

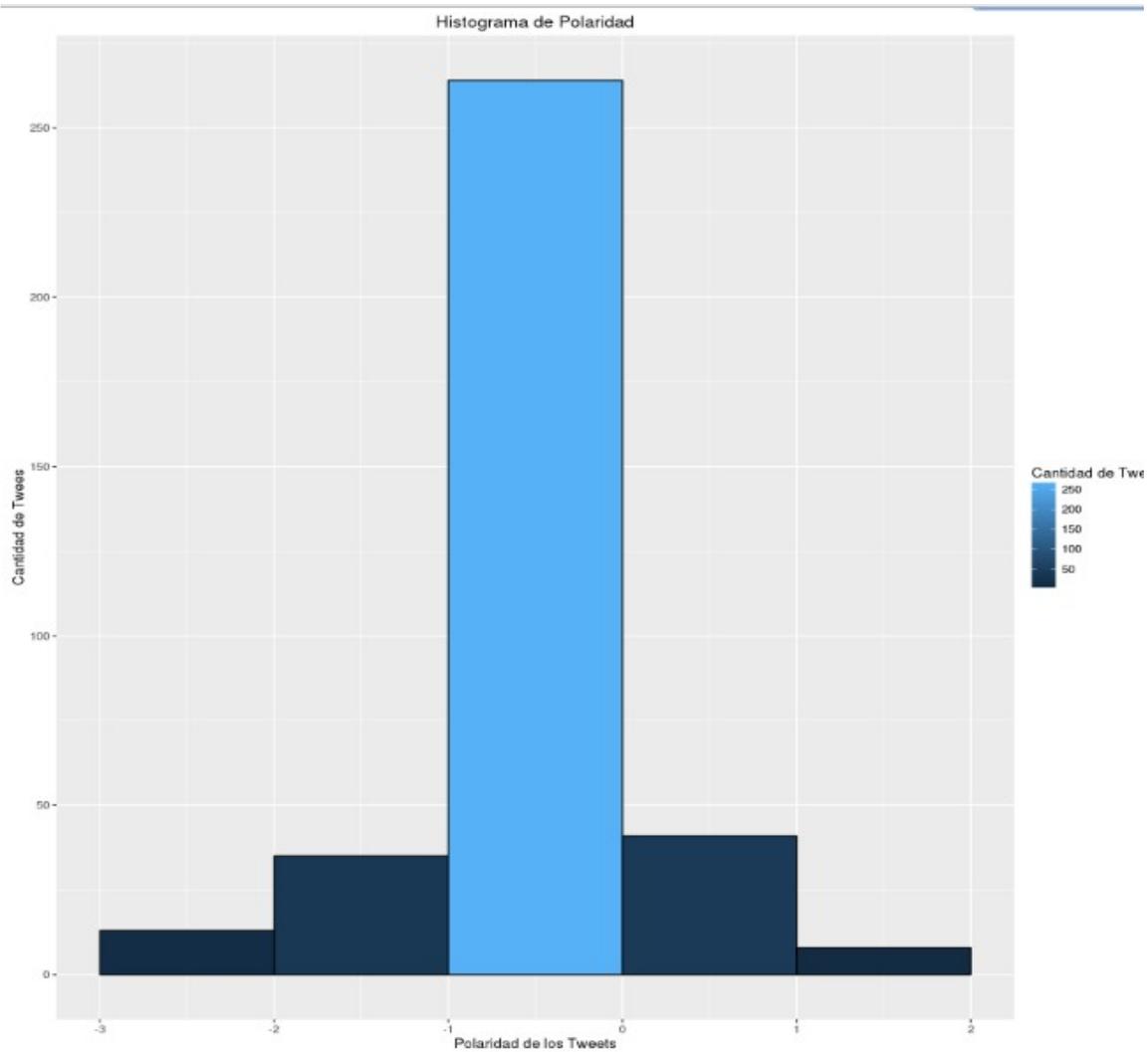


Figura 32. Histograma de polaridad

N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Carga
1	rivas_davila	545	UNICA	2016-10

Archivos de Limpieza Utilizados

N°	Archivo de Limpieza de Palabras	Cantidad de Palabras
1	español	308

Información de los Usuarios

N°	Nombre Usuario	Tweets por Usuario	Porcentaje de Tweets por Usuario
1	mauroQTA	1	0.18
2	Francis970AM	3	0.55
3	bebeisisnono	1	0.18
4	MetroNicarag	1	0.18
5	YamalDris	1	0.18
6	totoclerch	1	0.18
7	bearami18	2	0.37
8	CNOPslp_	1	0.18
9	lcocoyoc	2	0.37
10	AlfredoARan	1	0.18
11	jeelbu77	1	0.18
12	LocutorAlejo	1	0.18
13	1000_am	1	0.18
14	Mamarbel	1	0.18
15	ManuelSaltos	1	0.18
16	CarlosTeran	2	0.37
17	infocracia	1	0.18
18	tytyguevara	1	0.18
19	enrisanror	1	0.18
20	Comandante0	1	0.18
21	fe8dc3e10bae	1	0.18
22	PerezAzkatun	1	0.18
23	CEBB_241	1	0.18
24	LuisCastiglio	1	0.18
25	Fudettrujillo	1	0.18
26	Antonio2717	1	0.18
27	eguerri_	1	0.18
28	NoticiasNava	1	0.18
29	Esperanza_PP	1	0.18
30	Leonelbric	1	0.18
31	mujemaravil	2	0.37
32	RTfujitrolls	1	0.18
33	HMariaVarga	1	0.18
34	ramjos2	1	0.18
35	EzeFrias	1	0.18
36	Jaimegarciara	1	0.18

Figura 33. Reporte nube de palabras

A	B	C	
471	460mile		10.
472	evguilera		10.
473	yhonifrensala		10.
474	misarro		10.
475	dayamithey2		10.
476	BloqueChavi		10.
477	claribelMR		10.

Información de las Palabras			
N°	Palabra	Menciones	%
1	ampliacion	97	20.
2	parroquia	44	10.
3	via	42	10.
4	mes	37	10.
5	carriles	36	10.
6	americas	35	10.
7	icvarela	35	10.
8	licita	35	10.
9	mop	35	10.
10	nuevos	35	10.
11	proyectos	35	10.
12	viales	35	10.
13	nuevas	34	10.
14	panchoarias	33	10.
15	viviendas	33	10.
16	dignas	32	10.
17	entregar	32	10.
18	luevesdevivie	32	10.
19	leoni	32	10.
20	llegamos	32	10.
21	maracaibo	32	10.
22	raul	32	10.
23	mas	28	10.
24	mcpio	22	10.
25	familias	21	10.
26	asi	19	10.
27	equipo	19	10.
28	hoy	17	10.
29	hogares	14	10.
30	patria	14	10.
31	edificados	13	10.
32	junto	13	10.
33	reciben	13	10.
34	comunicacio	12	10.
35	doctrina	12	10.
36	gmyy	12	10.
37	iniciamos	12	10.
38	liderazgo	12	10.

Figura 33.1. Reporte nube de palabras

N°	Nombre Búsqueda	Items Encontrados	Fecha
1	rivas davila	545	UI

Información de los Usuarios			
N°	Nombre Usuario	Items por Usuario	Fecha
1	mauroQTA	1	10.0
2	Francis970A	3	10.0
3	bebeisisnono	1	10.0
4	MetroNicarag	1	10.0
5	YamalDris	1	10.0
6	totoclerch	1	10.0
7	bearami18	2	10.0
8	CNOPslp_	1	10.0
9	lcocoyoc	2	10.0
10	AlfredoARan	1	10.0
11	jeelbu77	1	10.0
12	LocutorAlejo	1	10.0
13	1000_am	1	10.0
14	Mamarbel	1	10.0
15	ManuelSalto	1	10.0
16	CarlosTeran	2	10.0
17	infocracia	1	10.0
18	tytyguevara	1	10.0
19	enrisanror	1	10.0
20	Comandante	1	10.0
21	fe8dc3e10bae	1	10.0
22	PerezAzkatun	1	10.0
23	CEBB_241	1	10.0
24	LuisCastiglio	1	10.0
25	Fudettrujillo	1	10.0
26	Antonio2717	1	10.0
27	eguerre_	1	10.0
28	NoticiasNava	1	10.0
29	Esperanza_P	1	10.0
30	Leonelbric	1	10.0
31	mujemaravil	2	10.0
32	RTfujitrolls	1	10.0
33	HMariaVanga	1	10.0
34	ramjos2	1	10.0
35	EzeFrias	1	10.0
36	Jaimegarciar	1	10.0
37	FSilvaFacetti	2	10.0
38	josemaferber	1	10.0
39	AshtrayGirl	1	10.0
40	cope_es	1	10.0
41	wakanoticias	1	10.0

Figura 34. Reporte Histograma de Polaridad

Información General de la Polaridad

N°	Polaridad	de Tweets por	Polaridad Respecto
1	-3	14	2.57
2	-2	23	4.22
3	-1	148	27.16
4	0	616	113.03
5	1	86	15.78
6	2	17	3.12
7	-6	1	0.18
8	-3	14	2.57
9	-2	23	4.22
10	-1	148	27.16
11	0	616	113.03
12	1	86	15.78
13	2	17	3.12
14	3	1	0.18

Tweets y sus Polaridades

N°	Texto del Tweet	Polaridad del Tweet
1	leoperiodista I	-3
2	leoperiodista I	-3
3	leoperiodista I	-3
4	leoperiodista I	-3
5	leoperiodista I	-3
6	leoperiodista I	-3
7	leoperiodista I	-3
8	leoperiodista I	-3
9	InformanteMF	-2
10	InformanteMF	-2
11	InformanteMF	-2

Figura 34.1. Reporte de Histograma de Polaridad

4.4 PRUEBAS

Las pruebas que se realizaron y que se mostraran a continuación están enfocadas a los proyectos del estado Mérida municipio Rivas Dávila.

Prueba 1: En la primera prueba se visualizan personas que comentan sobre diversos temas, no se consiguió información relevante a los proyectos realizados en Mérida municipio Rivas Dávila ya que la información fue muy general. Las palabras que se utilizaron para la búsqueda fueron las siguientes:

ampliación, puente , tapias, municipio, rivas, davila, merida, problema, cola,

Figura 36. Prueba1 Reporte Nube

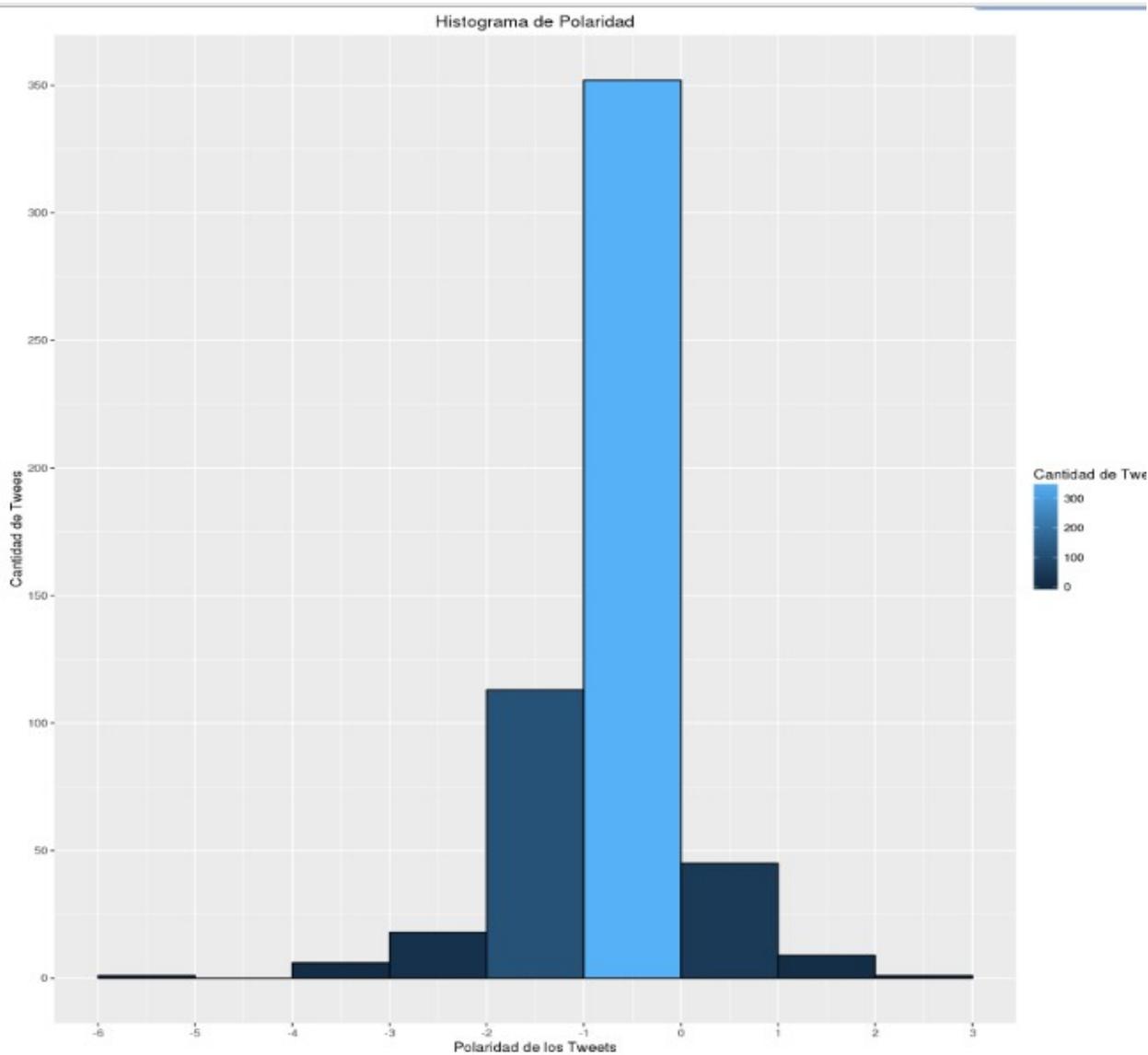


Figura 37. Prueba1 Histograma

A	B	C	D	E	F	G	H
Información de las Búsquedas							
N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Carga en el servidor	Fecha de Ejecución		
1	rivas davila	545	UNICA	2016-10-06 18:27:22			
Archivos de Limpieza Utilizados							
N°	Archivo de Limpieza	Cantidad de Palabras					
1	español	308					
Información de los Usuarios							
N°	Nombre Usuario	Tweets por Usuario	de Tweets por Usuario				
1	mauroQTA	1	0.18				
2	Francis970A	3	0.55				
3	bebeisisnono	1	0.18				
4	MetroNicarag	1	0.18				
5	YamalDris	1	0.18				
6	totoclerch	1	0.18				
7	bearami18	2	0.37				
8	CNOPslp_	1	0.18				
9	lcocoyoc	2	0.37				
10	AlfredoARan	1	0.18				
11	jeelbu77	1	0.18				
12	LocutorAlejo	1	0.18				
13	1000_am	1	0.18				
14	Mamarbel	1	0.18				
15	ManuelSalto	1	0.18				
16	CarlosTeran	2	0.37				
17	infocracia	1	0.18				
18	tytyguevara	1	0.18				
19	enrisanror	1	0.18				
20	Comandante	1	0.18				

Figura 38. Prueba1 Reporte Histograma

[3] " RT copees AMPLIACIÓN EL Parlamento catalán aprueba un referéndum en septiembre de 2017 "

[4] " AMPLIACIÓN EL aeropuerto de Fort Lauderdale cerró el jueves por la mañana "

[5] " RT GobiernodeCeuta AMPLIACIÓN] La Cuenta General de 2015 confirma la buena senda de la Hacienda de la Ciudad de Ceuta "

[6] " RT ElRolfiPy Senadores rechazan ampliación para pago de gratificaciones "

[7] " Universidad Nacional Itapua obtuvo su ampliación de 2374286000 Gs Pasa al Poder Ejecutivo "

[8] " RT aurelionuno La ampliación de la cobertura educativa es un logro del siglo XX del que debemos sentirnos orgullosos Senador "

[9] " RT lineacaliente Vecinos de Tepoztlán exige se reinicie la ampliación de la autopista La Pera Cuautla "

[10] " Hoy realizamos el riego diario de la ampliación de la Av Los Horcones y Av Jardines del Aeropuerto CiudadAbierta "

[11] " RT noabraspaz Comienza la ampliación del congreso RajoySi nos condenan por la Gürtel y hay terceras elecciones no habrá suf "

[12] " Encuentran un muerto en la Col Ampliación SANTA Lucía en una bolsa Río Las rosas X la técnica 139 en Oaxaca "

[13] " radio970AM Senado acepta ampliación presupuestaria para gobernación de Concepción y Misiones "

[14] " RT bearami18 Universidad Nacional Canindeyu obtuvo su ampliación de 454142950 Gs Pasa al Poder Ejecutivo 1000am "

[15] " RT copees AMPLIACIÓN EL Parlamento catalán aprueba un referéndum en septiembre de 2017 "

[16] " RT IESSec En noviembre se terminará la readecuación y ampliación del área de emergencia para brindar atención a 15191 benef "

[17] " Universidad Nacional Canindeyu obtuvo su ampliación de 454142950 Gs Pasa al Poder Ejecutivo 1000am "

[18] " AMPLIACIÓN 700 Hás por confirmar es la superficie afectada en incendio forestal en Placilla Valparaiso ViveCASABLANCA Bomba "

[19] " RT reformanacional Toman caseta de Tepoztlán manifestantes a favor de ampliación vial hay paso libre pero afecta tránsito "

[20] " RT leandroquerido Ml artículo sobre el Plebiscito en Colombia Ampliación de derechos políticos vs sobre representación abus "

[21] " AMPLIACIÓN 700 Hás por confirmar es la superficie afectada en incendio forestal en Placilla Valparaiso reddeemergencia infor "

[22] " RT JuanOrlandoH Contenido de ver la ampliación del CEB Miguel Andoníe Fernández en la colonia La Era "

[23] " RT copees AMPLIACIÓN EL Parlamento catalán aprueba un referéndum en septiembre de 2017 "

[24] " RT lineacaliente Se manifiestan en caseta habitantes de Tepoztlán exigen se continúe con la ampliación de la autopista La P "

[25] " RT Antequerajose FrustraronLaPazConMentiras Y ahora quieren liquidar el corazón del acuerdo la ampliación de la democracia "

[26] " NoticiasNavarra AMPLIACIÓN EL Parlament aprueba celebrar un referéndum en 2017 con o sin el aval del Estado "

[27] " RT CEBioBio QuillonAmpliacion se trata de colisión de dos vehículos un lesionado inmovilizado por bomberos Cat8 ^104 "

[28] " radio970AM Senado acepta ampliación presupuestaria para universidad de Itapua canindeyu concepción y rechazan para universi "

[29] " RT ElRolfiPy Senadores rechazan ampliación para pago de gratificaciones "

[30] " RT Rangelsilva2012 1 Entre las obras a las cuales pasamos revista están la ampliación del LB Mesa de Los Morenos en la parro "

[31] " PSOESVQ Espada no a la Ampliación de zona azul en Virgen del Rocío usted prometió quitar la zona azul si no es usted un embi "

[32] " RT SeniatRNO PrensaALDía EU Anuncian ampliación de capacidad productiva "

[33] " AMPLIACIÓN EL Parlament aprueba celebrar un referéndum en 2017 con o sin el aval del Estado "

[34] " RT A3Noticias AMPLIACIÓN EL repunte del desempleo en septiembre el segundo consecutivo es inferior al de 2015 GRÁFICO "

[35] " RT Rangelsilva2012 1 Entre las obras a las cuales pasamos revista están la ampliación del LB Mesa de Los Morenos en la parro "

[36] " RT bearami18 Senado rechaza ampliación presupuestaria de 25 mil millones Gs para Corte Suprema Justicia 1000am "

[37] " RT bearami18 Senado rechaza ampliación presupuestaria de 13 mil millones de Gs para Ministerio Público 1000am "

[38] " RT padmejones Desde las 11 pm del sábado 8 MuniLima cierra tránsito en vladucto San Borja Norte por ampliación del Derby so "

[39] " SecSatudGBZ PanchoArias2012 Comenzaron los trabajos de ampliación y remodelación de sala de parto y maternidad Tenta que se "

[40] " A favor del la pera Cuautla y el progreso que puede traer Morelos "

[41] " EzeFrias Ampliación "

[42] " RT Antequerajose FrustraronLaPazConMentiras Y ahora quieren liquidar el corazón del acuerdo la ampliación de la democracia "

[43] " RT ElRolfiPy Senadores rechazan ampliación para pago de gratificaciones "

[44] " RT copees AMPLIACIÓN EL Parlamento catalán aprueba un referéndum en septiembre de 2017 "

[45] " RT bosquenixticuil BoletínDenuncia ante PRODEURJal por ampliación ilegal de Av ÁngelLeaño que amenaza bosque Nixticuil "

[46] " AMPLIACIÓN EL Parlamento catalán aprueba un referéndum en septiembre de 2017 "

[47] " Néstor Reverol anunció la posible ampliación de los cuadrantes para fortalecer la seguridad "

[48] " Gobierno prevé nueva ampliación del Canal de Panamá "

[49] " Entonces es como si la dependencia pidiera un segundo periodo de ampliación porque obvio siempre amplía el plazo de respues "

[50] " RT bearami18 Senado rechaza ampliación presupuestaria de 13 mil millones de Gs para Ministerio Público 1000am "

[51] " sofíleonori Mostearnos mutuamente es como el puente hidrógeno que une nuestra amistad "

[52] " RT noticiaseul ML presupuesto firma con que el BCE celebra congreso y reformas "

Figura 39. Prueba1 Texto Plano

Prueba 2: En esta segunda prueba se visualizan personas que comentan sobre el proyecto de Mérida puente los rastros, nos muestra algo más concreto .Se realizaron combinación de palabras lo cual nos dio un resultado más certero. Las palabras que se utilizaron para la búsqueda fueron las siguientes:

rivas+davila+merida,geronimo+maldonado,libertador+merida,obispo+ramos+loras,santos+marquina,mucupiz+justo+briceño,parroquia+chiguara,chiguara+merida,obispo+ramos,marquin,mcupiz.

Y los resultados que mostraron se visualizaran en las siguientes figuras:

Información de las Búsquedas					
N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Carga en el Servidor	Fecha
1	rivas_davila_5	342	UNICA	2016-10-06 16:00	2016-10-06 16:00

Archivos de Limpieza Utilizados		
N°	Archivo de Limpieza	Cantidad de Palabras
1	español	308

Información de los Usuarios			
N°	Nombre Usuario	Tweets por Usuario	Calidad de Tweets por Usuario
1	pasatfc	3	0.88
2	PartidoADMe	2	0.58
3	FrankGavid1a	2	0.58
4	ADCulturaMe	2	0.58
5	ciemunidad	3	0.88
6	elvisdurant	2	0.58
7	GuidoMercado	2	0.58
8	RAFAEL_P1	2	0.58
9	Sara_Hernandez	2	0.58
10	JAD_Merida	2	0.58
11	JADCampoEl	2	0.58
12	AbilioGomez	2	0.58
13	NairDavila	2	0.58

Figura 41. Prueba2 Reporte Nube.

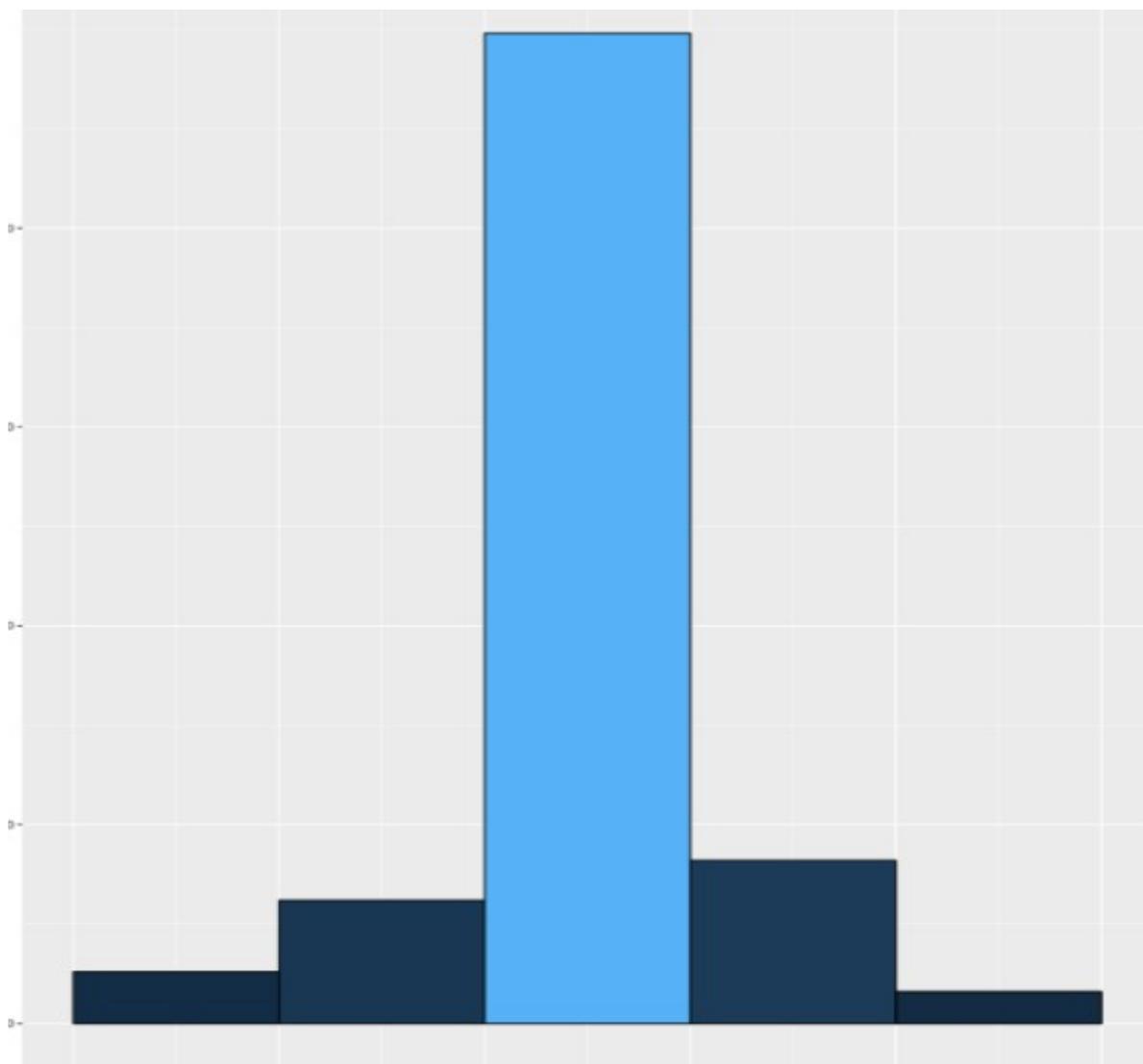


Figura 42. Prueba2 Histograma de Polaridad

Información de las Búsquedas					
N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Carga en el servidor	Fecha de Ejecución
1	rivas_davila	342	UNICA	2016-10-06 1	2016-10

Información de los Usuarios			
N°	Nombre Usuario	Tweets por Usuario	Calidad de Tweets por Usuario
1	pasatfc	3	0.88
2	PartidoADMe	2	0.58
3	FrankGavid1a	2	0.58
4	ADCulturaMe	2	0.58
5	ciemunidad	3	0.88
6	elvisdurant	2	0.58
7	GuidoMercad	2	0.58
8	RAFAEL_P1	2	0.58
9	Sara_Hernand	2	0.58
10	JAD_Merida	2	0.58
11	JADCampoEl	2	0.58
12	AbilioGomez	2	0.58
13	NairDavila	2	0.58
14	olga_eze	4	1.17
15	SumateMerid	3	0.88

Figura 43. Prueba2 Reporte Histograma de Polaridad

```

to (2) x
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" RT AbllioGomezZULA Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" Asi iniciamos con los talleres Doctrina Liderazgo y Comunicación junto al equipo AD Mcpio Rivas Dávila Merida "
" Retweeted Nair Dávila Rivas NairDávila Merida Ejido Te esperamos hoy viernes 30 EL PODER DEL PENSAMIENTO "
" Recolección de semillas de trigo en la Esc Estadal Boliv Las Playitas Mcpio Rivas Davila Merida NicolasMaduro "
" Para el 20 MP Rivas Dávila del EstadoMérida tendrá 1 centrorecepcióndeMV ubicado en Liceo DrGerónimo Maldonado con 4 pto de recepción
" RT INMIVI GobAlexisR construcción puente Los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1 "
" Siembra de plántulas de lechuga y ají dulce UE Geronimo Maldonado munc Ribas Davila MPEDUCACION NicolasMaduro "
" RT INMIVI GobAlexisR construcción puente Los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1 "
" GobdeMerida FONHIVIM RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorge
" FONHIVIM RT INMIVI GobAlexisR construcción puente Los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1 "
" RT INMIVI GobAlexisR construcción puente Los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1 "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT INMIVI GobAlexisR construcción puente Los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1 "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPIO RIVAS DÁVILA dcabellor "
" RT Mujeresjpsvumr La ineptitud del alcalde de libertador en Mérida ocasiona el segundo accidente en el municipio CarlosGarcíaInepto htt
" La ineptitud del alcalde de libertador en Mérida ocasiona el segundo accidente en el municipio CarlosGarcíaInepto "
" Charla CLAP Casa de Alimentación 1400044 Libertador propuestas de cursos NicolasMaduro RMarcoTorres "
" INN continúa con el Regreso a clases Mcpio Libertador en la EB Rafael Antonio Godoy innvzla NicolasMaduro "
" La cámara municipal de libertador organizo actividades para celebrar el 458 aniversario de la fundación de la ciudad de mérida "
" RT Djosfer Vladimiralaiyv politicaenVladimiralai porfa lee mi participación Aquí en merida libertador no llega los alimentos metante "
" Vladimiralaiyv politicaenVladimiralai porfa lee mi participación Aquí en merida libertador no llega los alimentos metante el ojo "
" Concejo Municipal del Libertador Mérida no tiene competencia sobre aumento del pasaje urbano "
" En Venta Terreno en Sector El Valle Mérida Mas información aquí gtgt 04147599244 "
" RT PjFamiliarNac Pancartazo en la parroquia Caracclolo Parra municipio Libertador Merida PrimeroJusticia "
" En Venta Terreno en Sector El Valle Mérida Mas información aquí gtgt 04147599244 "
" RT DoctsnvaEsparta FVM y Docentes Municipales exigen la homologación de sueldos y pago de deudas a la Alcaldía Libertador Mérida "
" En Venta Casa Quinta en Urb La Mata de Mérida Mas información aquí gtgt 04147599244 "
" En Venta Casa Quinta en Urb La Mata de Mérida Mas información aquí gtgt 04147599244 "
" Los Concejales del Libertador informaron sobre la programación del aniversario de Mérida "
" RT AdelainePo Los Concejales del Libertador dieron a conocer la programación para este 9 de octubre aniversario de merida "
" RT AdelainePo Los Concejales del Libertador dieron a conocer la programación para este 9 de octubre aniversario de merida "
" RT AdelainePo Los Concejales del Libertador dieron a conocer la programación para este 9 de octubre aniversario de merida "
" Los Concejales del Libertador dieron a conocer la programación para este 9 de octubre aniversario de merida "
" Concejales del Libertador informaron sobre la programación del aniversario de Mérida "

```

Figura 44. Prueba2 Texto Plano.

Prueba 3: Por último se realizó una tercera prueba, la cual quitamos palabras de la búsqueda anterior que vimos no relevantes y dejamos una, para ser más precisa dicha búsqueda, buscando tweets referentes al proyecto, basados en los resultados de los anteriores. En esta prueba se refleja la molestia por parte de algunos usuarios de Twitter por no culminar el puente los rastrojos que está en construcción y a la vez se puede ver que otras personas y el Alcalde publican que va en la segunda etapa la construcción de dicho puente . Las palabras que se utilizaron fueron las siguientes:

puente+rastrojos

Y los resultados que mostraron se visualizaran en las siguientes imágenes:



Figura 45. Prueba3 Nube de Palabra

N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Fecha Carga en el Archivo
1	puente rastrojo	19	UNICA	2016-10-07 11:00

Archivos de Limpieza Utilizados

N°	Archivo de Limpieza	Cantidad de Palabras
1	español	308

Información de los Usuarios

N°	Nombre Usuario	Tweets por Usuario	Calidad de Tweets por Usuario
1	MariaE1850	4	21.05
2	comtrerasaqu	1	5.26
3	Chlqul	2	10.53
4	rdfajardo	1	5.26
5	GobdeMerida	1	5.26
6	FONHVIM	1	5.26
7	Mocusa3	1	5.26
8	InjuvemRango	2	10.53
9	Fernand5846	1	5.26
10	JoseJai34491	1	5.26
11	460mile	1	5.26
12	evguilera	1	5.26
13	nadasirve16	1	5.26
14	ARAGUAPS	1	5.26

Información de las Palabras

N°	Palabra	Menciones	Calidad de Palabras por Tweets
1	construccion	14	7.14
2	etapa	14	7.14

Hoja 1

Figura 46. Prueba3 Reporte Nube

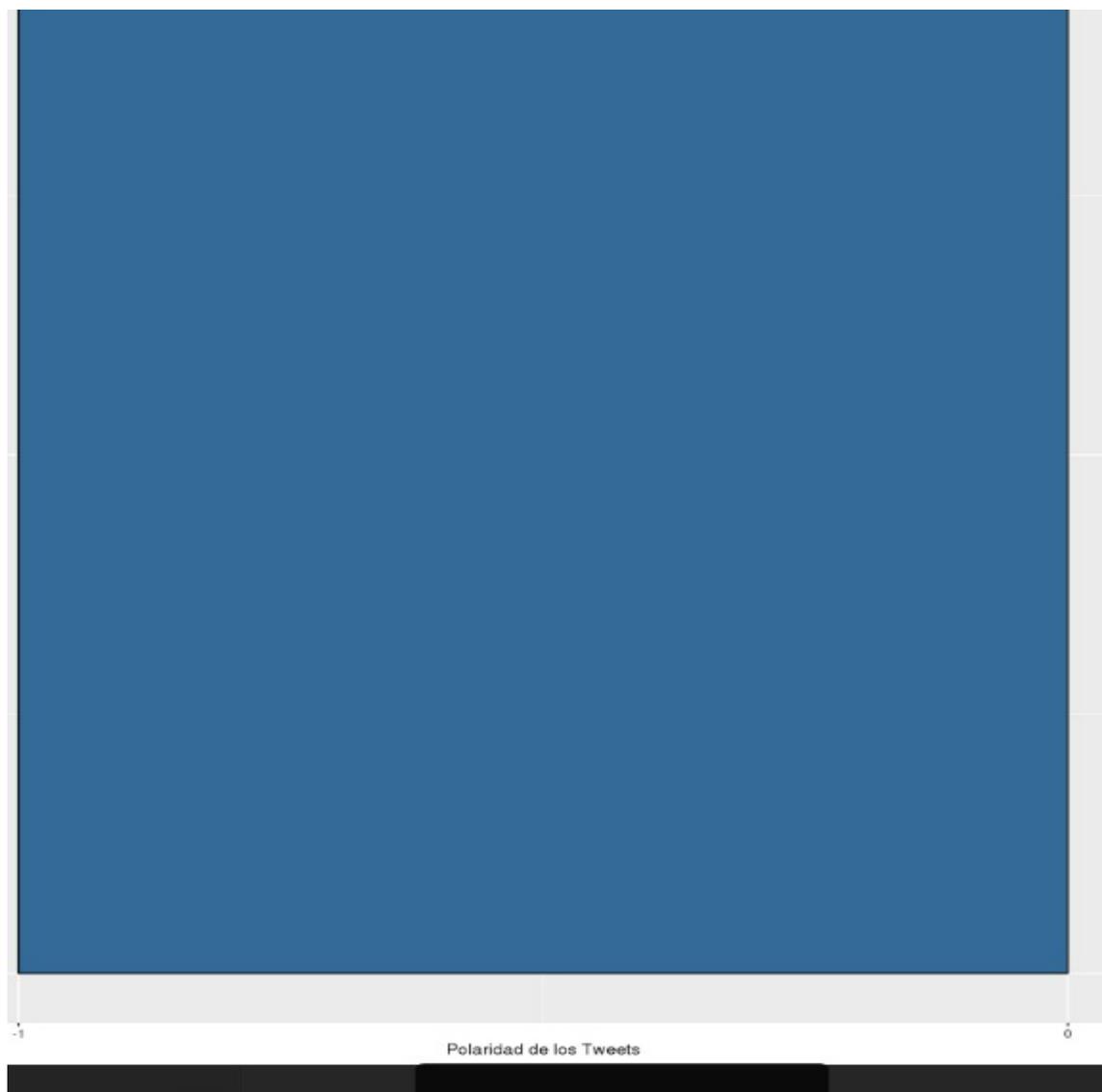


Figura 47 Prueba3 Histograma de Polaridad

A	B	C	D	E	
Información de las Búsquedas					
N°	Nombre Búsqueda	Tweets Encontrados	Calidad Búsqueda	Carga en el servidor	Fecha
1	puente rastrojo	19	UNICA	2016-10-07 1	2016-1
Información de los Usuarios					
N°	Nombre Usuario	Tweets por Usuario	Calidad de Tweets por Usuario		
1	MariaE1850	4	21.05		
2	comtrerasraqu	1	5.26		
3	Chlqul	2	10.53		
4	rdfajardo	1	5.26		
5	GobdeMerida	1	5.26		
6	FONHVIM	1	5.26		
7	Mocusa3	1	5.26		
8	InjuvemRange	2	10.53		
9	Fernand58461	1	5.26		
10	JoseJai344913	1	5.26		
11	460mile	1	5.26		
12	evguilera	1	5.26		
13	nadasirve16	1	5.26		
14	ARAGUAPSU	1	5.26		
Información General de la Polaridad de los Tweets					
N°	Polaridad	Calidad de Tweets por Polaridad	Calidad Respecto al Total de Tweets		
1	-1	4	21.05		
2	0	15	78.95		
Tweets y sus Polaridades					
N°	Texto del Tweet	Polaridad del Tweet			
1	josbarrerasBu	-1			
2	josbarrerasBu	-1			
3	madrugonazo	-1			

Figura 48. Prueba3 Reporte Histograma

```

:to (3) ×
" josbarreras Buenos días al Alcalde Barrera se le agradece que terminen el puente de los Rastrojos a Atapaima dejaron la cal
" josbarreras Buenos días al Alcalde Barrera se le agradece que terminen el puente de los Rastrojos a Atapaima dejaron la cal
" madrugonazo1 Buenos días al Alcalde Barrera se le agradece que terminen el puente de los Rastrojos a Atapaima dejaron la cal
" madrugonazo1 Buenos días al Alcalde Barrera se le agradece que terminen el puente de los Rastrojos a Atapaima dejaron la cal
" RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1
" RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1
" RT GobAlexisR Retweeted INMIVIMERIDA INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo "
" GobdeMerida FONHVIM RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas I
" FONHVIM RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorge
" RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPPIO RIVAS DÁVILA dcabellor
" RT GobAlexisR Retweeted INMIVIMERIDA INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo "
" RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPPIO RIVAS DÁVILA dcabellor
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPPIO RIVAS DÁVILA dcabellor
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPPIO RIVAS DÁVILA dcabellor
" RT ANDREZ77 GobAlexisR CONSTRUCCIÓN PUENTE LOS RASTROJOS II ETAPA PARROQUIA GERONIMO MALDONADO MCPPIO RIVAS DÁVILA dcabellor
" josbarreras barreras cuando arreglaras de una vez por todas el paso del puente situado despues de la Y de los rastrojos haci
" RT INMIVI GobAlexisR construcción puente los rastrojos II Etapa Parroquia Geronimo Maldonado McpioRivas Davila Jorgebecerra1
" josbarrerasbuenos días al alcalde barrera se le agradece que terminen el puente de los rastrojos a atapaima dejaron la calle
" josbarrerasbuenos días al alcalde barrera se le agradece que terminen el puente de los rastrojos a atapaima dejaron la calle
" madrugonazo buenos días al alcalde barrera se le agradece que terminen el puente de los rastrojos a atapaima dejaron la calle
" madrugonazo buenos días al alcalde barrera se le agradece que terminen el puente de los rastrojos a atapaima dejaron la calle
" rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas davila jorgebecerra"
" rt gobalexisr retweeted inmivimerida inmivigobalexisr construccion puente los rastrojos ii etapa parroquia geronimo"
" gobdemerida fonhvim rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas d
" fonhvim rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas davila jorgeb
" rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas davila jorgebecerra"
" rt andrez gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcpio rivas davila dcabellor"
" rt gobalexisr retweeted inmivimerida inmivigobalexisr construccion puente los rastrojos ii etapa parroquia geronimo"
" rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas davila jorgebecerra"
" rt andrez gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcpio rivas davila dcabellor"
" rt andrez gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcpio rivas davila dcabellor"
" rt andrez gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcpio rivas davila dcabellor"
" rt andrez gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcpio rivas davila dcabellor"
" josbarreras barreras cuando arreglaras de una vez por todas el paso del puente situado despues de la y de los rastrojos hacia
" rt inmivi gobalexisr construccion puente los rastrojos ii etapa parroquia geronimo maldonado mcporivas davila jorgebecerra"

```

Figura 49. Prueba3 Texto plano

CAPITULO 5

5.1. CONCLUSIONES

Con todo lo antes mencionado, se deja constancia del correcto funcionamiento del sistema, cumpliendo con cada uno de los objetivos específicos planteados:

- Levantamiento de requerimientos, nos reunimos con los usuarios finales, los cuales explicaron cuál era su necesidad, así llegamos a la pregunta que englobaba el problema "¿Cómo extraer nuevas variables para la toma de decisiones a partir del análisis de redes sociales como twitter?". Una vez conociendo el problema, se abordó proponiendo un sistema web que realice análisis sobre la precepción social de los proyectos financiados por el Consejo Federal de Gobierno a partir de los contenidos publicados en Twitter; seguidamente se propuso una serie de técnicas y sistemas existentes disponibles que facilitara la elaboración de dicho sistema tales como, análisis de sentimientos, minería de texto, minería de redes sociales, entre otros. A su vez se propuso la base de datos que se utilizaría, en este caso se escogió Mongo DB por ser una base de datos que permite el manejo de grandes cantidades de datos, puesto que un previo análisis se observó que se trabajaría con una gran cantidad de información..
- Diseñar la interfaz, en el sistema desarrollado se buscó crear interfaces intuitivas permitiendo al usuario final realizar los procesos de búsqueda, limpieza, análisis y ver resultados. Dicho objetivo fue logrado con éxito.
- Crear roles y usuarios, se crearon dos roles que pueden tener los usuarios, analistas o administradores, con sus respectivas funcionalidades.
- Realizar extracción, limpieza y procesamiento, primero la información es extraída desde Twitter según la selección del usuario; segundo, a estos tweets se les aplica un proceso de limpieza removiendo enlaces, números, signos de puntuación, espacios en blanco, palabras que no aportan información relevante al análisis, caracteres extraños y transformando todas las palabras a minúsculas; por

último se realizan procesos de agrupación, conteo y transformación de información para obtener resultados.

- Mostrar resultados del procesamiento de la información recopilada, la información se muestra de forma gráfica y en forma de reportes en hojas de cálculo para que el usuario pueda interpretarla fácilmente.

Cumpliendo con cada uno de estos objetivos específicos podemos concluir que cumplimos el objetivo general del trabajo especial de grado "Desarrollar una aplicación web para realizar análisis de percepción social por medio de Twitter para su uso en el Consejo Federal de Gobierno".

5.2. CONTRIBUCIÓN

Este trabajo de investigación contribuye principalmente al Consejo Federal de Gobierno, ya que esta aplicación fue creada inicialmente para ser utilizada por los analistas y gerentes de esta institución. Más allá del Consejo Federal de Gobierno, esta aplicación puede ser útil para todo aquel que desee realizar un análisis de sentimientos por medio de las redes sociales específicamente Twitter y así conocer el impacto que puede tener algún producto, persona, institución, entre otros, ya sea positivo, negativo o neutro por medio de las redes sociales.

5.3. RECOMENDACIONES

Para ejecutar el sistema creado es necesario que el usuario tengan claro las palabras que estén relacionadas a los proyectos específicamente de vialidad para que así la búsqueda que desea realizar tenga un mejor resultado.

5.4. TRABAJOS FUTUROS

Sobre el trabajo realizado existen distintas modificaciones que pueden ser hechas para mejorar aún más la aplicación. Entre ellas están:

- Aplicar un framework para que el código de la aplicación sea más óptimo ya que actualmente está realizado en PHP puro.
- Poder conectarse a las API de otras redes sociales y así poder realizar análisis de sentimiento no solo con Twitter.
- Agregar búsquedas por defectos relacionadas tanto a vialidad como a otras categorías de los proyectos.
- Agregar la opción de utilizar el API streaming para trabajar con análisis de tweets en tiempo real.
- Crear otro rol para que solo pueda realizar las búsquedas. Y así el analista solo se encargue de ver resultados.
- Crear búsquedas por grupo para así tomar un conjunto de las búsquedas anteriores y ver de la combinación de los grupos, que resultados trae.

REFERENCIAS BIBLIOGRÁFICAS Y DIGITALES.

- [1] Ley N° 5963E. Ley Orgánica Del Consejo Federal De Gobierno, Caracas, Venezuela, 22 de febrero de 2010.
- [2] Java, A., Song, X., Finin, T. & Tseng, B.(2007).Why we twitter: understanding.
- [3] Jara, J., Nuñez,M. & Pezzino,S.(2013). Análisis de sentimientos de tweets. Universidad Católica “Nuestra Señora de la Asunción”, Paraguay.
- [4] Dev.twitter.com.(2016).Oauth | Twitter Developers. <https://dev.twitter.com/oauth>.
- [5] Dev.twitter.com.(2016).The Streaming APIs | Twitter Developers. <https://dev.twitter.com/rest/public/rate-limiting>.
- [6] MongoDB.(2016).NoSQL Databases Explained. <https://www.mongodb.com/nosql-explained>.
- [7] Lopez,D.(2016). Bases de Datos No Relacionales[Diapositiva en PowerPoint]. <http://es.slideshare.net/dipina/bases-de-datos-no-relacionales-nosql>.
- [8] Garcia, H. & Yanes, O.(2012) Bases de datos NoSQL. Telem@tica, 11(3). Recuperado desde: <http://revistatelematica.cujae.edu.cu/index.php/tele/article/view/74/74>.
- [9] Antiñanco, M.(2013).Bases de Datos NoSQL: Escalabilidad y alta disponibilidad a través de patrones de diseño. Universidad Nacional de La Plata, Argentina.
- [10] Castro, A., González, J. & Callejas, M.(2012). Utilidad y funcionamiento de las bases de datos NoSQL. Revista Facultad de Ingeniería - UPTC. 21(33). Recuperado desde: <https://dialnet.unirioja.es/servlet/articulo?codigo=5029469>.
- [11] Bustos, L.(2014).Análisis de sentimientos en Twitter. Universidad Nacional Autónoma de México, México.
- [12] Documents.software.dell.com.(2016).Text Mining-Statistics Textbook. <http://documents.software.dell.com/Statistics/Textbook/Text-Mining>.
- [13] Zafarani, R., Abbasi, M. & Liu, H., (2014), Social Media Mining An Introduction, New York USA, Cambridge University Press.
- [14] Reputacion Online, Netnografia & ARS.(2014). Social Media Mining – Reputacion Online,Netnografia & ARS.

- <http://migueldefresno.com/2014/10/social-media-mining-o-cuando-la-herramienta-demonitorizacion-no-es-la-clave.html>.
- [15] Charu, C. & Chandan, K., (2014). Data Clustering Algorithms and Applications, New York USA, CRC Press.
- [16] Anguita, M., Lorenzo, R., (2014). Extracción, análisis y visualización de información social desde Twitter. Universidad Complutense de Madrid, España.
- [17] Spina, V., (2014). Entity-Based Filtering And Topic Detection For Online Reputation Monitoring In Twitter. Universidad Nacional De Educación A Distancia, España.
- [18] Castellanos, A., Cigarrán, J. & García, A., (2014). Using IR Techniques for topic-based sentiment analysis through divergence models. Universidad Nacional De Educación A Distancia, España.
- [19] Pete Chapman, Julian Clinton, Randy Kerber, Thomas Khabaza, Thomas Reinartz, Colin Shearer y Rüdiger Wirth (2000). "CRISP-DM 1.0". URL: <http://the-modeling-agency.com/crisp-dm.pdf>.
- [20] A. Alliance, "Guide to agile practices," <http://guide.agilealliance.org/>, 2013, [Online; Accedido el 19-Julio-2015].
- [21] K. Beck, Extreme Programming Explained, 2da. ed. Addison-Wesley, 1999.